Issue No.18 PRACTICAL MATHEMATICS Volume No.3

ISSUE No.18

PRACTICAL MATHEMATICS

A QUARTERLY

WITH APPLICATIONS IN THE FIELD OF BUSINESS

MATHEMATICS IN STATISTICAL METHOD

Central Tendency and Dispersion
Mean • Median • Mode
Frequencies • Ogives
Kurtosis • Range • Skewness
Deviation

Time Series
Secular Trend • Cyclical Variation

Sampling • Index Numbers
Standard Error

Functional Relationships
Regression • Correlation

Index to Volume III

REGINALD STEVENS KIMBALL

50¢

EDITOR: REGINALD STEVENS KIMBALL ED.D.

ISSUE 18

Practical Mathematics

VOLUME 3

REGINALD STEVENS KIMBALL, *Editor*

TABLE OF CONTENTS

PRACTICAL MATHEMATICS is published quarterly by the National Educational Alliance, Inc., Office of publication at Washington and South Avenues, Dunellen, N. J. Executive and editorial offices, 37 W. 47th St., New York 19, N. Y. John J. Crawley, President; A. R. Mahony, Vice President and Business Manager; Frank P. Crawley, Treasurer. Issue No. 18, August 20, 1944. Entered as second-class matter April 10, 1943, at the Post Office at Dunellen, N. J., under the Act of March 3, 1879. Printed in the U.S.A. Price in the U.S.A., 50c a copy; annual subscription at the rate of 50c a copy.

CHATS WITH THE EDITOR

PRACTICAL MATHEMATICS presents for its eighteenth issue a group of articles on business statistics. In recent years, statistical analysis has come to play an important part in the business world and the man or woman who is able to prepare and present reports based on correct statistical procedure is much in demand in all kinds of enterprises.

Much of the matter considered in statistical method is simply mere common sense. Figures may be manipulated in many ways, all of them mathematically correct, but not all of them leading to equally valid results. Many a business has "gone on the rocks" simply because the figures employed in analyzing conditions were used incorrectly and hence produced misleading results.

Many promotional schemes are so dressed up with "statistics" that they appear to tell a story which is far different from the truth of the matter. Figures don't lie, it is certain, but liars do figure. Not all false reports are out-and-out "lies" in the sense that the compiler of the figures has deliberately set out to present a false picture of conditions. Not infrequently, he deceives himself as well as his readers.

In the small scope of one issue, it is impossible for PRACTICAL MATHEMATICS to do more than sketch the part which mathematics plays in the field of statistical method. To gain a complete understanding of the subject, the reader should refer to any one of the number of excellent books which are on the market. The areas covered in this issue are those in which mathematics—principally arithmetic, with a dash of algebra and coördinates—plays a principal part. Throughout the discussion, we have suggested by means of cross references that the reader turn back to the various sections of earlier issues of PRACTICAL MATHEMATICS where a fuller treatment of certain processes has been given. This serves a double purpose: it conserves space and at the same time stimulates a review of the earlier issues.

We begin our study of statistical method with a consideration of a rather simple but typical example. Throughout the first section, we submit the same figures to several different treatments and then compare the results. In most cases, we have used actual figures taken from some field of business or industry in our illustrative examples. In a few cases, in order to simplify the computations and put the stress on the acquisition of facility in using the method, we have deliberately "manufactured" the data. In each case, the reader is apprised of the source of the figures so that he may know whether they are drawn from reality or have been prepared to help teach a principle.

In those instances where it has been deemed necessary to use technical terms (so that the reader may be able to converse intelligently on the subject and to follow the treatment of standard works in the field of statistical analysis), we have defined the term directly in the text.

In discussing measures of central tendency, we have made a quick review of the discussion about the average which appeared in the numbers of PRACTICAL MATHEMATICS

devoted to arithmetic and then have gone on to the presentation of additional measures which have not previously been introduced. The chart (p. 1202) showing the utility of each of these measures, as well as any handicaps which may result from their use, is worthy of considerable study. Familiarity with this chart will be of great assistance in helping the reader to avoid many of the pitfalls which are ever present in the use of statistics.

In the section on measures of dispersion, we have occasion to make mention of the Gaussian curve, or curve of normal distribution, which has come to be one of the important stand-bys in all statistical analysis. It should be recognized that the hypothetical normal curve is interpreted as being the exact record of an infinite number of cases. In any sample, no matter how large, we realize that we are merely attempting to judge the whole field by means of the cases which we have under view at a given moment. We do not expect that every computation will result in a normal curve; if that were to be expected, we should not need to go through the computational processes which we employ. What we are interested in doing is to see how the curves in which our computations result compare with the assumed normal curve.

Throughout the discussion, the reader must keep in mind that his computations will lead to results which are only approximate. Each answer is subject to a certain degree of correction because of "error". The use of the word, *error*, in statistical work may need a bit of explanation. By "error" in this sense, we do not mean that there is a *mistake* in the computations—that we have added 2 and 2 to get 5, or something of that sort. Rather, we mean that the data have not been sufficient to give us an exact picture, but that we may make use of our results provided we "take them with a grain of salt". For that reason, when stating results, we always give the reader an idea of the degree of accuracy of the figures. The reader who trains himself to do this regularly will find that he is consciously looking for the correction factor when he reads statistical findings presented in the newspapers or in other sources from which the general public derives its information.

The section on time series and variation shows one method by which the trend of developments in any field may be judged. Indeed, it is possible by means of these formulas to predict the probable future development of an industry. The only caution to be observed is to remember that unforeseen factors may crop up at a later date to nullify to some extent the usual course of events. Particularly in a time of great uncertainty such as the present wartime period, there is much likelihood that day-to-day developments will be such that predictions based on present trends may not actually come to pass.

For instance, a business built up in a boom town will undoubtedly suffer reverses if later a substantial proportion of the population moves away from the town in which the business is located; hence, any figures based on the present growth and development of that business must be considered as of a temporary nature, with consequent revision at a later date of the findings based upon them.

The section on sampling gives a brief explanation of the methods by which such nation-wide investigations as the Gallup Poll are carried on. Usually, each investigator develops his own formulas on which to base his findings, the exact nature depending on the type of information which is being sought. The principles underlying all of them are similar, however, and one who has mastered this section will have a better basis by which to determine the extent of credence

which he will give such reports. Some investigations of this nature flourish for a time and then vanish from sight because the public loses faith in them after discovering that they have been based on false premises; others continue for relatively long periods because their originators have hit upon a remarkably accurate formula. Here, again, common sense as well as mathematics is brought into play.

We have devoted only a few pages to the matter of preparing index numbers, largely because the present is not an appropriate time for making much use of them. The uncertainties of war-time conditions are such that many of the index reports which have been current for a long period of years have been suspended. Our discussion of index numbers and our illustrations of the manner in which several are derived is sufficient to show the reader the general principles involved and to indicate the extent to which they are credible. The reader who plots for himself the curves based on any two or three of them will discover that the curves present a considerable degree of similarity even though the figures appear to be wholly dissimilar. We have not gone into the matter of interpreting index numbers because whole volumes have been devoted to this subject and even the makers of such numbers are not entirely agreed among themselves on all points of settling validity.

As we approach the article on functional relationships, we engage in a bit of fascinating by-play. Here we introduce the subject with a coin toss that presents opportunities for amplification through untold hours of repetition. By this simple little device, we open to the reader vistas of more or less amusing and profitable investigations.

The subject of correlation is one with which everyone working in the field of statistical method should be familiar. The many different forms of correlation offer interesting opportunities for further investigation after one had become familiar with the general principles. Here, again, we have limited our discussion to the bare mathematical elements. The reader who is desirous of knowing more about the applications of correlation to specific problems is referred to the complete treatises on the subject.

To a greater extent than any of its predecessors, this issue of PRACTICAL MATHEMATICS is besprinkled with formulas. It is not necessary that all of these be committed to memory. Rather, it is a better procedure to become familiar with their use and then to turn to the pages on which they are presented when it is desired to apply any one of them in an investigation. To assist the reader, we have grouped them in one spot (pp. 1267-1269), so that they may more readily be brought into use when occasion demands.

The many tables which are presented in the final section of the issue will be found of great help in performing the computations which are demanded in the employment of the various formulas. To work all of the examples out long-hand would take countless hours which might be devoted to better purpose. In this connection, the reader should not forget the tables of common logarithms (pp. 126, 127), which will be of use in easily attaining the multiplications and divisions called for by a number of the formulas. If you want to do one or two of the problems in each group the "hard way" for purposes of finding out by comparison the extent to which the tables give you sufficient accuracy, well and good. From that point on, we advise you to make use of the tables and to "round off" your figures so that you may save your time and energy.

On pages 1270 and 1271, the reader will find tables of the natural, or

Naperian, logarithms. All of the work in PRACTICAL MATHEMATICS has been based on common, or Briggan, logarithms, but we have had occasion to mention the Naperian system several times (pp. 219, 256, 690) and many of our readers have requested that we present the tables of natural logarithms in order that they might be able readily to compare the two systems. The fact that the natural logarithms are printed in this issue, which is devoted to statistical method, should not mislead the reader into thinking that natural logarithms are to be preferred to common logarithms for statistical purposes. They are given here simply to help the reader round out his concepts of logarithmic devices.

For nearly two years, a staff of workers in the offices of the National Educational Alliance and more than thirty professors on the campuses of almost as many colleges and universities have been engaged in distilling out of the whole realm of mathematics those items which are essential to a practical working knowledge of the subject. We have concentrated our attention on those subjects which are uppermost in the minds of men and women who are today engaged in carrying on war activities and in planning for the peace which we hope is not too far distant. Every problem has met the scrutiny of a number of experts; special assistance has been employed in solving, checking, and re-checking the solutions.

To the countless numbers of readers who have taken the trouble to write their suggestions to the editor, we are all greatly indebted. Their efforts have enabled us better to estimate the reception which PRACTICAL MATHEMATICS has been accorded throughout the United States and even in far-flung portions of the world. Not a few of our readers are actually engaged in personal combat, as is attested by the number of V-mail letters which have come into the editorial office. Some are men and women well along in life; others are just beginning their careers. Their continued interest in this magazine throughout the period of its existence has been a constant stimulus to all of us engaged in the enterprise to put our best efforts into helping them to learn what mathematics is and what it can do for them.

To them, as well as to the hundreds of students in his own classes who have served as "guinea pigs" in the stages when the manuscript was in preparation, the editor offers his personal thanks. The opportunity to try out some of the suggestions offered in these pages and to observe at first hand the extent to which they were readily understandable has greatly facilitated the preparation of the entire work.

As a final word of caution, at the risk of repetition, the editor wishes once again to emphasize some of the suggestions which he has made in his earlier chats. You can't learn mathematics simply by placing the book under your pillow. A certain amount of hard, conscientious, persistent practice is necessary to the mastery of any subject. As you find yourself growing "rusty" on any process (and all of us do if we don't keep everlastingly at it), turn back to the place where that process is explained and go over it step by step. You'll be surprised to discover how fast it all "comes back" to you. Above all, don't neglect the fundamental operations of arithmetic. As we have warned you over and over again, ninety per cent of all mistakes in mathematics are due to failure to get these simple number combinations set down correctly. If the foundations are faulty, nothing based on them can stand.

R. S. K.

· MATHEMATICS IN STATISTICAL METHOD ·

By Reginald Stevens Kimball, Ed. D.

BEFORE numerals—whether they represent counts or measurements —can be treated mathematically, they must be arranged in some sort of order so that they can be made to tell a story. The same set of figures can be interpreted in several different ways, with the result that several different stories may be read from them. All of these stories may be *computationally* correct, yet not all of them, strictly speaking, may be *mathematically* correct. It is our purpose, in the present article, to attempt to determine the correct method of analyzing figures to attain a defensible and reliable result.

PUTTING FIGURES TO WORK

In order to get a general idea of what "statistics" or "statistical method" is all about, let us first consider a typical problem of the sort frequently encountered in the business world. As we subject the figures to various interpretations, we shall be developing some of the concepts which we shall then study more formally in the later sections of this article.

Illustrative Example

The sales tickets for a certain department in a large store showed amounts for a given day as listed in Table 1.

TABLE 1

SALES-TICKET TOTALS—IN ORDER OF OCCURRENCE

2.45	2.28	2.72	1.83	1.75	2.69	3.25	3.17	2.69	2.72	2.55	3.11
2.28	1.16	2.45	2.79	2.28	2.21	2.50	2.13	2.99	2.21	3.17	2.55
3.49	2.31	1.73	1.89	2.50	1.11	2.21	2.45	2.50	2.25	2.50	2.79
1.61	2.11	3.20	1.81	2.45	2.55	1.81	3.75	2.13	2.87	2.25	2.87
2.69	1.80	3.19	2.50	2.11	2.31	1.95	3.11	3.49	3.17	2.55	2.50
3.89	1.55	2.50	2.01	3.84	3.39	1.25	2.89	2.28	2.45	2.79	3.05
3.73	1.48	2.13	2.55	2.31	2.28	2.45	1.95	3.60	3.20	2.69	2.99
2.50	2.89	2.01	3.25	2.25	1.51	2.87	2.31	2.25	2.89	3.11	2.55
2.72	1.27	2.50	1.51	1.75	1.89	2.45	1.80	2.45	3.25	3.05	2.87
2.99	3.45	1.75	1.40	2.50	2.01	2.13	3.52	2.75	2.69	2.31	2.69
1.25	2.72	2.50	2.50	2.69	2.72	3.27	2.31	2.11	2.55	2.75	1.83
1.95	2.31	2.55	2.75	1.89	2.79	3.75	1.83	3.19	2.21	3.05	2.75

Taken as they stand, these figures present only a confusing jumble. We are not able to determine at a glance the trend of the day's sales. We might arrive at this trend by totalling the items and taking the average (p. 61),

$$\$360.00 \div 144 = \$2.50,$$

but this involves a rather lengthy addition which we desire to avoid if we can find another way of discovering what we want to know.

Orderly array

One way of doing this is to re-arrange our figures in order of size, from the smallest to the largest. In this case, the figures would appear as in Table 2.

TABLE 2

SALES-TICKET TOTALS—IN RISING ORDER OF VALUE

1.11	1.75	1.89	2.13	2.28	2.45	2.50	2.55	2.72	2.87	3.11	3.27
1.16	1.75	1.95	2.13	2.28	2.45	2.50	2.55	2.72	2.87	3.11	3.39
1.25	1.75	1.95	2.21	2.28	2.45	2.50	2.69	2.75	2.89	3.17	3.45
1.25	1.80	1.95	2.21	2.31	2.45	2.50	2.69	2.75	2.89	3.17	3.49
1.27	1.80	1.95	2.21	2.31	2.45	2.50	2.69	2.75	2.89	3.17	3.49
1.40	1.81	2.01	2.25	2.31	2.45	2.50	2.69	2.75	2.89	3.19	3.52
1.48	1.81	2.01	2.25	2.31	2.50	2.55	2.69	2.79	2.99	3.19	3.60
1.51	1.83	2.11	2.25	2.31	2.50	2.55	2.69	2.79	2.99	3.20	3.73
1.51	1.83	2.11	2.25	2.31	2.50	2.55	2.69	2.79	3.05	3.20	3.75
1.55	1.83	2.11	2.25	2.31	2.50	2.55	2.72	2.79	3.05	3.25	3.75
1.61	1.89	2.13	2.28	2.45	2.50	2.55	2.72	2.87	3.05	3.25	3.84
1.73	1.89	2.13	2.28	2.45	2.50	2.55	2.72	2.87	3.11	3.25	3.89

We may now readily see that the largest sale was $3.89 and the smallest $1.11. Almost as easily, we see that the mid-point in the list is $2.50 and that there are more figures for sales of $2.50 than for any other amount.

Tallying

Rather than re-copy all of these figures in this fashion, however, we could more easily have prepared a distribution table which would have given us these same facts and several more. To do this, we could make use of the tally system as a labor-saving device.

The tally system consists of drawing a straight line for each occurrence, with a diagonal line drawn through four straight lines to indicate the fifth.

Thus, ~~||||~~ ~~||||~~ || would indicate twelve (12). Its advantage lies in the fact that we have merely to count the groups of 5's instead of having to count a large number of individual lines. When using this system, we first arrange our numbers in a column at the left and then, taking the numbers in the order of their appearance in Table 1, draw the check-mark for each in turn. Thus, we are sure that we have not skipped any numbers in making our tally.

In ordinary practice, we should probably use every figure, beginning with what appears to be the smallest, so that our complete tally would look like the small portion reproduced in Table 3. A firm daily engaged in checking its accounts in this fashion would probably have printed or mimeographed sheets of numbers to save the tedious task of drawing up a long numerical list of this sort. In our reproduction of

TABLE 3

PARTIAL TALLY OF SALES TICKETS

Price	Tally	Occurrences
1.10		
1.11	/	1
1.12		
1.13		
1.14		
1.15		
1.16	/	1
1.17		
1.18		
1.19		
1.20		
1.21		
1.22		
1.23		
1.24		
1.25	//	2
1.26		
1.27	/	1

the tally-table for the figures from Table 1, we shall, for the purpose of conserving space, omit numbers which did not actually occur in the sales of the day under consideration. Table 4 shows the final figures as thus determined.

These same facts might be portrayed graphically as in Fig. 1 (p. 1188), where the interval between two consecutive lines on the *X*-axis represents twenty-five cents ($0.25) and the interval between two lines on the *Y*-axis represents one (1) sale.

From these facts, we readily determine that the *mode* (the figure appearing most frequently) is $2.50. By a count of the number of cases (144), we find that the *median** (the middle case) is also $2.50.

Class intervals

In arranging frequency distributions, we get a better picture, ordinarily, by grouping the numbers in class intervals, taking any convenient range. For the figures which we have under consideration at present, we might take intervals of twenty-five cents ($0.25), counting twelve and one-half cents ($0.125) either way from the even twenty-five cent figures, so that the even quarters would represent the mid-values of each range. The distribution of tallies, under this arrangement, would take the form shown in Table 5. This is the form in which most concerns would keep their tally sheets, rather than in the form shown as an example of preparation in Table 4.

TABLE 4

COMPLETE TALLY OF SALES TICKETS

Price	Tally	Occurrences
1.11	I	1
1.16	I	1
1.25	II	2
1.27	I	1
1.40	I	1
1.48	I	1
1.51	II	2
1.55	I	1
1.61	I	1
1.73	I	1
1.75	III	3
1.80	II	2
1.81	II	2
1.83	III	3
1.89	III	3
1.95	III	3
2.01	III	3
2.11	III	3
2.13	IIII	4
2.21	IIII	4
2.25	IIII	4
2.28	~~IIII~~	5
2.31	~~IIII~~ II	7
2.45	~~IIII~~ III	8
2.50	~~IIII~~ ~~IIII~~ II	12
2.55	~~IIII~~ III	8
2.69	~~IIII~~ II	7
2.72	~~IIII~~	5
2.75	IIII	4
2.79	IIII	4
2.87	IIII	4
2.89	III	3
2.99	III	3
3.05	III	3
3.11	III	3
3.17	III	3
3.19	II	2
3.20	II	2
3.25	III	3
3.27	I	1
3.39	I	1
3.45	I	1
3.49	II	2
3.52	I	1
3.60	I	1
3.73	I	1
3.75	II	2
3.84	I	1
3.89	I	1

Following the same plan that was used on page 232 in developing a bar graph, we may, by placing our rectangular bars one next the other, develop a *histogram*, or *column diagram*, for these figures, as in Fig. 2. Here we see clearly that the bulk of the sales—roughly one-third of the total—are found in the class intervals whose mid-points extend from $2.25 to $2.75, with the number of sales below $2.25 and above $2.75 tapering off sharply.

* The median for 144 cases is found by adding 1 to 144, getting 145, and then dividing by 2, the result being 72.5. Since both the seventy-second and the seventy-third cases (cf. Table 2) are $2.50, we do not have to interpolate to find the value of the median lying between these two cases.

TABLE 5

TALLY OF SALES TICKETS BY CLASS INTERVALS

Class Interval i	Mid-value m	Tallies	Frequency f
0.875-1.125*	1.00	I	1
1.125-1.375	1.25	IIII	4
1.375-1.625	1.50	~~IIII~~ I	6
1.625-1.875	1.75	~~IIII~~ ~~IIII~~ I	11
1.875-2.125	2.00	~~IIII~~ ~~IIII~~ II	12
2.125-2.375	2.25	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ IIII	24
2.375-2.625	2.50	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ III	28
2.625-2.875	2.75	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ IIII	24
2.875-3.125	3.00	~~IIII~~ ~~IIII~~ II	12
3.125-3.375	3.25	~~IIII~~ ~~IIII~~ I	11
3.375-3.625	3.50	~~IIII~~ I	6
3.625-3.875	3.75	IIII	4
3.875-4.125	4.00	I	1

If now we connect the mid-points of the tops of the bars in this histogram by straight lines, we get a curve, as shown in Fig. 3. This shows us roughly the extent of the sales and at the same time tends to give us a more intelligible picture by eliminating the sharp corners of the rectangular bars. The same curve is repeated in Fig. 4 (without the bars). In this form, it is known as a *curve of distribution* or *frequency polygon*. In this instance, we discover that, if we fold the curve along a line dropped from its mid-point to the base-line, the two halves of the curve coincide exactly.

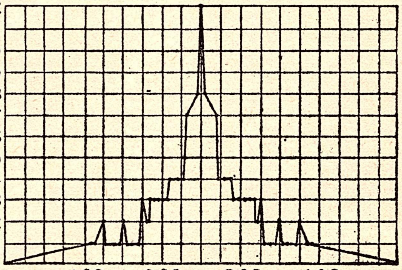


Fig. 1

In other words, the curve is symmetrical. This is a good test of the reliability we may place upon our interpretation of the figures. In instances where the curve is not symmetrical, as we shall see in later examples, we need to consider whether some of the figures are so atypical or unusual that they distort the figure and handicap us in making our interpretations.

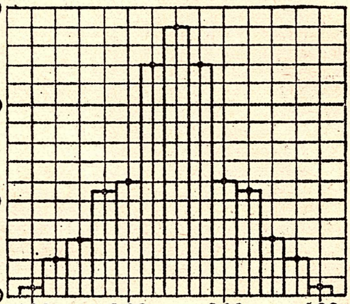


Fig. 2

A normal curve of distribution, taking into account an infinite number of cases, would follow the form shown in Fig. 5. This curve†, named for the mathematician who determined the formula on which it is based, is also

*It should be noted here that the actual class interval is really interpreted as stopping just short of 1.125 (say at 1.12499999 . . .) so that we may avoid the question as to where to include an item that is apparently at the upper extreme of one class and at the lower extreme of the next higher class.

†The formula for this curve is $y = \frac{n}{\sigma\sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}}$.

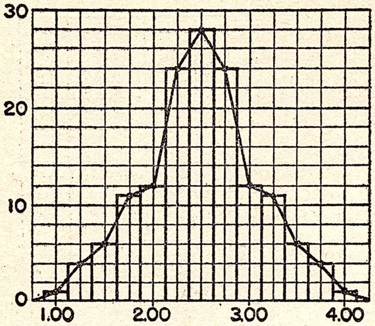


Fig. 3

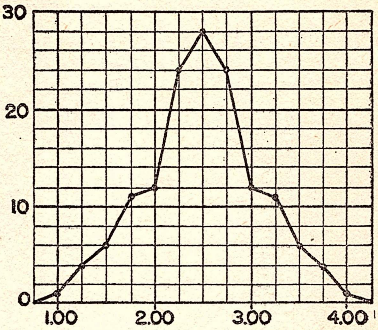


Fig. 4

known as a Gaussian curve. The values of the ordinates of this curve are detailed in Table XCIX, page 1255.

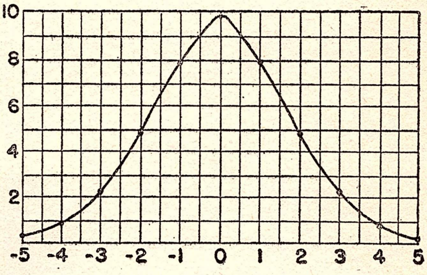


Fig. 5

Ogives

A further picture of the sales situation which we have been considering may be achieved by finding the *ogives*. An ogive is found by tabulating the cumulative frequencies; that is, as we come to each class interval, we add to its frequency all of the frequencies which have preceded. Cumulated frequencies may be reckoned from either end of the distribution range. Thus, we have *more-than* ogives and *less-than* ogives, as shown in Figs. 6 to 8.

To prepare the ogive, we pick up the final figure in the first column and the number in the "frequency" column already presented in Table 5, adding another column for the cumulated frequencies.

From this, for instance, we see that 34 sales were for less than $2.125. In the same way, we may just as easily determine the number of sales

TABLE 6

CUMULATED FREQUENCIES OF SALES—RISING ORDER

Less Than	Frequency	Cumulated Frequency
1.125	1	1
1.375	4	5
1.625	6	11
1.875	11	22
2.215	12	34
2.375	24	58
2.625	28	86
2.875	24	110
3.125	12	122
3.375	11	133
3.625	6	139
3.875	4	143
4.125	1	144

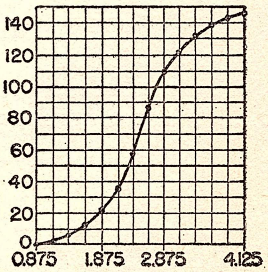


Fig. 6

TABLE 7

CUMULATED FREQUENCIES OF SALES—DESCENDING ORDER

More than	Frequency	Cumulated Frequency
3.875	1	1
3.625	4	5
3.375	6	11
3.125	11	22
2.875	12	34
2.625	24	58
2.375	28	86
2.125	24	110
1.875	12	122
1.625	11	133
1.375	6	139
1.125	4	143
0.875	1	144

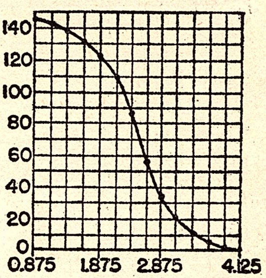


Fig. 7

under any specified amount simply by reading the number which appears in the cumulated frequency column opposite the figure with which we are concerned. This ogive would be plotted as in Fig. 6.

To prepare the more-than ogive, we reverse the order of cumulation. In Table 7, the pertinent material in Table 5 is copied from bottom to top, this time taking the initial figure from the "class interval" column and the numbers in the "frequency column". In actual practice, however, the actual figures are not reversed. The more-than ogive for this table is shown in Fig. 7.

By plotting the two ogives on the same graph, as in Fig. 8, we can ascertain the mid-point of the sales values without resorting to the numerical computation which we performed on page 1187. The point at which the two ogives cross (in this case $2.50) can be read directly from the graph—72 sales above $2.50 and 72 sales below that figure.

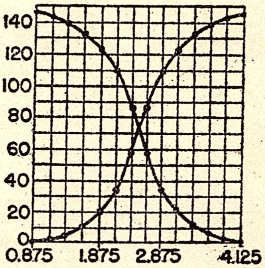


Fig. 8

The case which we have been considering thus far has involved a symmetrical distribution, as we noted on page 1188 when we folded the curve along the axis of its central ordinate. Most of the figures with which we shall have to deal in actual business affairs are not quite so accommodating.

Let us next consider a fairly simple case in which the curve is not symmetrical. This time, we shall group all of our tabulations in one table, in order that we may more readily make comparisons. In reading this table, you will want to take into account what has previously been said in connection with the preparation of the different parts.

Illustrative Example

The number of units of work completed each week by the various workers in a certain factory was found to be as shown in the third column of Table 8.

TABLE 8

WEEKLY OUTPUT IN TERMS OF COMPLETED UNITS

Number of Pieces	Tallies	Frequency	Cumulated Frequencies Less Than	Cumulated Frequencies More Than	Totals
n		f	$<$	$>$	fn
464	𝍸	5	5	57	2320
465	𝍸 |	6	11	52	2790
466	𝍸 |	6	17	46	2796
467	𝍸 ||	7	24	40	3269
468	𝍸 |||	8	32	33	3744
469	𝍸 ||||	9	41	25	4221
470	𝍸 ||	7	48	16	3290
471	𝍸 |	6	54	9	2826
472	|||	3	57	3	1416
Totals		$\Sigma f = 57$			$\Sigma fn = 26672$

Mode = 469 Arithmetic Mean = 26672 ÷ 57 = 467.9 Median = $\frac{57+1}{2}$th = 29th = 468

We find the mode directly from Table 8, at the point, 469, where there is the greatest number of tallies. The mean may be computed by adding the figures in the "totals" column (which gives the product of each number in the first column multiplied by the corresponding number in the third column and dividing by the total number of cases, the total of the third column. The median, the value of the $\frac{n+1}{2}$th case, here the $\frac{57+1}{2}$th or the 29th case, would fall, as determined from either ogive column, somewhere near 468.

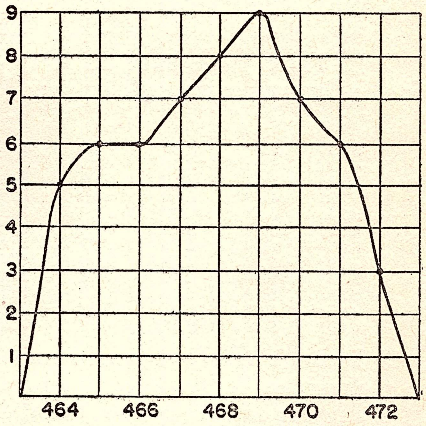


Fig. 9

Plotting the distribution curve, as in Fig. 9, we note that the peak is to the right of the central ordinate, but that the right half of the curve slopes more sharply than does the left half. These facts help us to understand why we do not get the exact correspondence of mean, median, and mode that we did in the preceding illustrative example. However, there is a fairly close correspondence in this case, so that we should not go far wrong in our consideration if we relied upon any one of the three as our guide.

In a case where a much greater divergence is discovered, one must proceed with caution in selecting any one of the three "averages" for further use. Fig. 17 (p. 1202) will be of great help in your analysis of these discrepancies.

At this point, the reader may begin to wonder why it is necessary to prepare graphs as well as to tabulate the figures which are being considered. In actual practice, one who is skilled in this work would, many times, not go to the labor of doing both, as he would be able to

visualize from the tabulated material the general appearance of the graph or to read from the graph a close approximation to the tabular values. Until one is highly familiar with the process, however, added understanding comes from performing both tasks for the same set of figures.

Having seen from this preliminary study of illustrative examples some of the procedures followed in putting figures to work, we are now ready to review the materials and to add to our knowledge of the terms employed.

TEST YOUR KNOWLEDGE OF FREQUENCY DISTRIBUTION

1 Table α presents the random figures of production as reported by foremen in a munitions factory. Tally these figures to show the frequency for each.

TABLE α

404	401	399	401	403	400	403	402	401	404
407	403	403	399	399	403	400	403	401	403
401	405	405	403	401	407	398	401	400	401
398	399	398	403	403	403	405	404	401	399
402	406	404	401	401	400	402	400	401	402
404	400	401	405	404	402	399	404	402	401
400	403	406	402	400	405	404	399	400	400
402	408	402	400	403	404	405	404	401	402
405	398	401	403	402	399	402	401	399	401
403	402	405	406	400	402	404	402	401	400
402	404	398	399	403	400	401	400	400	402
406	402	403	402	405	404	401	402	401	401

2 In Table β, we have an array of sales-ticket totals similar to that used in the illustrative example on page 1185. Arrange and tally by class intervals.

TABLE β

5.15	5.29	5.49	4.73	4.92	5.34	4.71	5.19	5.38	5.19
4.59	5.10	5.22	4.45	5.22	5.29	5.15	5.25	5.29	5.81
5.17	5.72	5.15	5.69	5.17	5.79	5.17	5.97	5.45	5.10
5.22	4.99	5.88	5.99	5.38	5.85	5.34	4.55	5.34	5.25
5.34	4.83	5.17	5.19	5.22	5.17	5.19	6.00	5.15	4.67
4.89	5.39	5.22	5.34	5.95	4.51	5.39	5.17	5.25	5.83
6.05	5.29	5.25	5.22	5.00	5.52	5.22	5.29	5.92	4.49
5.19	5.22	5.38	5.10	5.75	5.22	4.69	5.19	5.39	5.75
5.29	5.89	5.17	5.25	5.57	5.15	5.12	5.25	5.25	5.34
5.38	5.19	5.61	5.22	5.12	4.57	5.29	4.79	4.95	4.61
4.75	5.22	5.25	5.38	5.22	5.12	4.47	5.34	5.50	5.15
4.65	5.39	5.12	4.63	5.25	5.25	5.29	5.17	4.86	5.19

3 Draw (a) the histograms and (b) the frequency polygons for the data presented in Tables α and β.

4 Prepare frequency tables and draw curves for them, based on the tables you have prepared in answer to problems **1** and **2**.

MEASURES OF CENTRAL TENDENCY

The best-known measure of central tendency, or *average*, is the one which we rather fully described in our early work in arithmetic (pp. 75ff.). Its special name, to distinguish it from other types of averages, is the *arithmetic mean*. It may be expressed algebraically as

$$M_A = \frac{\Sigma v}{n} \qquad \text{Ia}$$

where Σ is used to indicate the process of adding,
v = the value of each individual item
and n = the number of items.

The process is shown on page 75.

For the weighted average, discussed on page 76, the formula is modified to take the weighting into account:

$$_{w}M_A = \frac{\Sigma(_{w}v)}{\Sigma w} \qquad \text{Ib}$$

where w is used to indicate the process of weighting.

The arithmetic mean in frequency distributions

In the case of a frequency distribution involving class intervals, we get an approximate average by taking

$$_{a}M_A = \frac{\Sigma(fm)}{n} \qquad \text{Ic}$$

or as

$$_{a}M_A = \frac{\Sigma(fm)}{\Sigma f}. \qquad \text{Id}$$

Returning to Table 5 for our example, we could here determine the arithmetic mean by multiplying each mid-value by its frequency, adding the products, and dividing by the number of cases as shown in Table 9.

Then $$_{a}M_A = \frac{360}{144} = 2.50.$$

TABLE 9

MID-VALUE COMPUTATION OF SALES TICKETS

Class Interval i	Mid-Value m	Frequency f	Product fm
0.875-1.125	1.00	1	1.00
1.125-1.375	1.25	4	5.00
1.375-1.625	1.50	6	9.00
1.625-1.875	1.75	11	19.25
1.875-2.125	2.00	12	24.00
2.125-2.375	2.25	24	54.00
2.375-2.625	2.50	28	70.00
2.625-2.875	2.75	24	66.00
2.875-3.125	3.00	12	36.00
3.125-3.375	3.25	11	35.75
3.375-3.625	3.50	6	21.00
3.625-3.875	3.75	4	15.00
3.875-4.125	4.00	1	4.00
		$\Sigma f = 144$	$\Sigma(fm) = 360.00$

The fact that we have taken mid-point values instead of the whole range of values does not change the total in the case of a symmetrical distribution because any excess at one point is compensated for by a like deficiency at another point, and *vice versa*. Even though we do not get exact compensation in some non-symmetrical cases, the greater ease of computation is an argument in favor of using this formula whenever the number of cases involved is extremely large.

The median

The median, the point on either side of which half the values in the series lie, is found by arranging the figures in an ascending or a descending order, counting the number of cases, adding one to the number thus obtained, and dividing by 2.

$$M_d = \frac{n+1}{2} \qquad \text{IIa}$$

Given the figures, 30, 35, 40, 45, 50, 55, and 60, representing the number of cents per hour paid to 7 different employees, we find that here $n=7$. Then

$$n+1=8 \text{ and } \frac{n+1}{2}=\frac{8}{2}=4.$$

The fourth item, counting from either end (in this instance, 45), represents the median wage.

If there is an even number of cases, as there would be if we eliminated the initial 30 from the list just utilized, we get a fractional point for a median. We should find here

$$n=6,\ n+1=7, \text{ and } \frac{n+1}{2}=\frac{7}{2}=3.5.$$

The third item from the left would be 45; the third from the right 50; the median would be half-way between the two numbers, at 47.5.

Sometimes we have to find the median by interpolation. Here we take $M_d=\frac{n}{2}$.

Let us take a case in which the employees of an organization are classified according to age, as shown in Table 10.

TABLE 10

AGES OF EMPLOYEES

Age Intervals	Mid-point	Number of Employees	Cumulated Frequencies Less Than	More Than
i	m	f	<	>
22.5-27.5	25	120	120	2013
27.5-32.5	30	152	272	1893
32.5-37.5	35	170	442	1741
37.5-42.5	40	214	656	1571
42.5-47.5	45	410	1066	1357
47.5-52.5	50	429	1495	947
52.5-57.5	55	259	1754	518
57.5-62.5	60	152	1906	259
62.5-67.5	65	107	2013	107
		$n=2013$		

Considering the number of cases, 2013, we find that

$$\frac{n}{2}=1006.5.$$

This figure would lie between 656 and 1066 in the "less than" column or between 947 and 1357 in the "more than" column. Both of these ranges are opposite the mid-point value, 45.

Using the less-than frequencies, we subtract the smaller number, 656, from both the larger number and the median.

$$\begin{array}{r} 1066 \\ -\ 656 \\ \hline 410 \end{array} \qquad\qquad \begin{array}{r} 1006.5 \\ -\ 656 \\ \hline 350.5 \end{array}$$

Thus, it is evident that the exact median lies $\frac{350.5}{10}$ of the way through the interval, or at

$$42.5+\frac{350.5}{410}\cdot 5=42.5+\frac{1752.5}{410}=42.5+4.27=46.77.$$

Had we used the more-than frequencies, our computations would have been similar; again we should have subtracted the smaller figure from both the larger figure and the median:

$$\begin{array}{r} 1357 \\ -\ 947 \\ \hline 410 \end{array} \qquad\qquad \begin{array}{r} 1006.5 \\ -\ 947 \\ \hline 59.5 \end{array}$$

$$47.5-\frac{59.5}{410}\cdot 5=47.5-\frac{297.5}{410}=47.5-0.73=46.77.$$

This procedure may be reduced to formula thus:

$$_iM_d=L_1+\frac{\frac{n}{2}-\Sigma_1}{f}\cdot i \qquad \textbf{IIb}$$

where L_1 = the lower limit of the class in which the median appears;
Σ_1 = cumulation of frequencies to lower limit of the median class;
f = the class frequency;
and i = the class interval,

or

$$_iM_d=L_2-\frac{\frac{n}{2}-\Sigma_2}{f}\cdot i \qquad \textbf{IIc}$$

where L_2 = the upper limit of the class in which the median appears;
and Σ_2 = cumulation of frequencies to upper limit of the median class.

The mode

The *modal average*, or the mode, is the term used to designate that figure which most frequently recurs in a group or series. When the number of cases under consideration is small, the mode is of little value as an index to the trend of the whole series. Its value as an index increases in proportion to the number of cases considered. When we are dealing with individual items, the mode is, of course, determined exactly. It is that one of the items which appears in the list most often, as we saw on page 1187 when considering Table 2.

When we are dealing with a frequency distribution expressed in class intervals, the mode cannot be determined with the same degree of exactness. We can approximate it, however, by an inspection of the frequencies in the various classes, using the formula,

$$_iM_o=L_m+\frac{f_h}{f_l+f_h}\cdot i \qquad \textbf{III}$$

where L_m = the lower limit of the frequency group which contains the mode;
f_h = the frequency of the class interval next higher in order than the modal group;
and f_l = the frequency of the class interval next lower in order than the modal group.

In the case of a "perfect distribution", such as the one we were considering in Table 5, we look for the largest number in the frequency column. Finding 28 opposite the class interval, 2.375-2.625, we select the lower limit, 2.375, as our L_m. The frequency of the next higher class, 24, is the figure we want for f_h, and the frequency of the next lower class, also 24, will be substituted for f_l.

Placing these values in our formula, we get

$$M_o = 2.375 + \frac{24}{24+24} \cdot 0.25 = 2.375 + \frac{24}{48} \cdot 0.25$$
$$= 2.375 + 0.5(0.25) = 2.375 + 0.125 = 2.50,$$

the exact figure which we have already determined by inspection. (We have gone to this labor here simply for the purpose of demonstrating the working of the formula.)

If we had been working with a group of figures in which the distribution was not perfectly symmetrical, we should not have obtained quite so accurate a result. Let us consider the situation shown in Table 11.

TABLE 11

ANNUAL SALARIES OF EMPLOYEES

Class Interval	Mid-point	Frequency	Product	Cumulated Frequencies Less Than	More than
i	m	f	fm	<	>
\$1150-1250	1200	4372	5246400	4372	54572
1250-1350	1300	3299	4288700	7671	50200
1350-1450	1400	3003	4204200	10674	46901
1450-1550	1500	3587	5380500	14261	43898
1550-1650	1600	4092	6547200	18353	40311
1650-1750	1700	11129	18919300	29482	36219
1750-1850	1800	18578	33440400	48060	25090
1850-1950	1900	3238	6152200	51298	6512
1950-2050	2000	1971	3942000	53269	3274
2050-2150	2100	1303	2736300	54572	1303
		$\Sigma f = 54572$	$\Sigma fm = 90857200$		

First, let us plot them as in Fig. 10. Then by formula III, since

$$L_m = 1750;$$
$$f_h = 3238;$$
$$f_l = 11129;$$

and $i = 100;$

$$M_o = \$1750 + \frac{3238}{11129+3238} \cdot 100$$
$$= \$1750 + \frac{323800}{14367}$$
$$= \$1750 + 22.53 = \$1772.53.$$

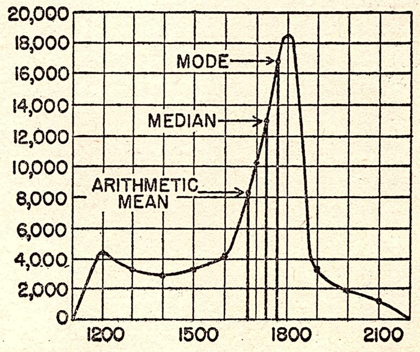


Fig. 10

Let us see how this compares with the median and the mean. Using formula Id and taking the values from Table 11, we have:

$$_iM_A = \frac{90857200}{54572} = \$1664.91.$$

To employ formula IIb in computing the median, we substitute these values, determined from the same table:

$$L_1=1650;\ \frac{n}{2}=27286;\ \Sigma_1=18353;\ f=11129;\ \text{and } i=100.$$

$$_iM_d=1650+\frac{27286-18353}{11129}\cdot 100=1650+\frac{893300}{11129}=1650+80.27=1730.27,$$

a number which agrees with neither the arithmetic mean nor the mode. Checking this last computation with formula IIc, where

$$L_2=1750 \quad \text{and} \quad \Sigma_2=25090,$$

we get

$$_iM_d=1750-\frac{27286-25090}{11129}\cdot 100=1750-\frac{219600}{11129}=1750-19.73=1730.27,$$

an exact correspondence with our previous computation. A plotting of the ogives, as in Fig. 11, gives further corroboration.

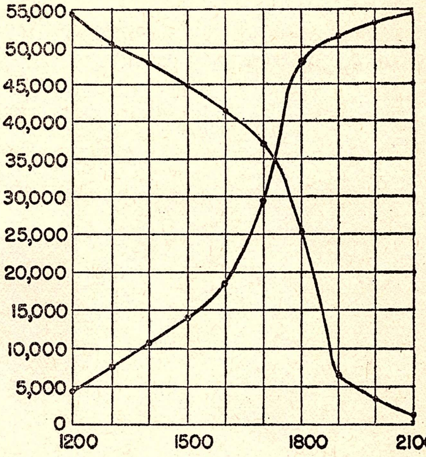


Fig. 11

This is a good illustration of a skewed distribution (see page 1212). When the distribution is skewed toward the the high values, we should always find

$$M_A<M_d<M_o,$$

whereas if it were skewed toward the low values, we should find exactly the opposite,

$$M_A>M_d>M_o.$$

Determination as to which of the three measures is the most significant can be made in any instance only after a close examination of the frequency tables to try to discover which factors are responsible for the distortion. The presence of a disproportionate number of items at either end of the scale will cause a skewing of the figures with the result that the "average" (the arithmetic mean) will be pulled in that direction to such an extent that it may not present a true picture of conditions. If the mode is found too near either end of the scale, it, too, will not be representative.

TEST YOUR KNOWLEDGE OF COMMON MEASURES OF CENTRAL TENDENCY

5 Compute the arithmetic mean for the data presented in Tables α and β.
6 Find the median for the figures shown in Tables α and β. Compare the results of your computation with the points at which the ogives meet in the graphs you prepared in answer to problem 4.
7 What is the mode for the data in Table α?
8 In the case of Table β, determine (a) the class interval in which the mode occurs; (b) the interpolated value of the mode.

The geometric mean

When we are concerned with averaging ratios or rates of change, we shall find the geometric mean a better guide than the arithmetic mean. It is found by obtaining the product of the items and then extracting the nth root of the product (n representing the number of items):

$$M_G=\sqrt[n]{a_1 \cdot a_2 \cdot a_3 \cdot \ldots \cdot a_n}. \qquad \text{IVa}$$

Does this sound like a particularly hard job? It would be if we had to perform the actual multiplications, but, fortunately, we can put logarithms to work for us. Remembering what we have already learned about using logarithms in order to avoid laborious multiplication (pp. 93–97) and for the purpose of extracting a root (p. 100), we may resolve the formula into a more usable form:

$$\log M_G=\frac{\log a_1+\log a_2+\log a_3+\ \ldots\ +\log a_n}{n}. \qquad \text{IVb}$$

In the case of a frequency distribution, this becomes:

$$\log {}_fM_G=\frac{f_1 \log m_1+f_2 \log m_2+f_3 \log m_3+\ \ldots\ +f_n \log m_n}{\Sigma f}. \qquad \text{IVc}$$

Table 12 shows the ratios of 86 companies, arranged according to class intervals.

TABLE 12

COMPARISON OF 86 COMPANY RATIOS—BY CLASS INTERVALS

Class Interval i	Mid-point m	Frequency f	fm	$\log m$	$f \log m$
40-50	45	3	135	1.65321	4.95963
50-60	55	14	770	1.74036	24.36504
60-70	65	17	1105	1.81291	30.81947
70-80	75	24	1800	1.87506	45.00144
80-90	85	15	1275	1.92942	28.94130
90-100	95	11	1045	1.97772	21.75492
100-110	105	2	210	2.02119	4.04238
		$\Sigma f=86$	$\Sigma(fm)=6340$		$\Sigma(f \log m)=159.88418$

Then $$\log {}_fM_G=\frac{159.88418}{86}=1.85912$$

$$ {}_fM_G=72.29,$$

whereas $$ {}_fM_A=\frac{6340}{86}=73.72.$$

In the case, the difference between the arithmetic mean and the geometric mean is scarcely sufficient to warrant the additional labor involved in computing the latter but there are occasions, as indicated on page 1202, when it is wiser to utilize this form of average.

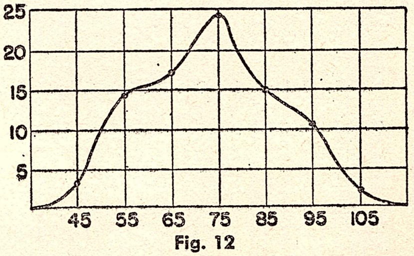


Fig. 12

Plotting the facts in Table 12 by the ordinary method gives us a curve such as that shown in Fig. 12. Since we are here concerned with ratios, however, it would

be more reasonable to use a logarithmic scale, by which we may more readily make comparisons as to the true relationships existing at all times between the items.

LOGARITHMIC GRAPHS

Logarithmic graph paper is ruled in accordance with the principles discussed in the introduction to logarithms (p. 90) and the slide rule (p. 102, paragraph 5). Semi-logarithmic graph paper has logarithmic rulings on one scale and arithmetic rulings on the other. Fig. 13 is a diagrammatic representation of the difference in the two rulings. In the upper line, the numbers are evenly spaced, after the fashion of the familiar "ruler" or yardstick. In the lower line, we have three logarithmic scales, the first extending from 1 to 10, the second from 10 to 100, and the third from 100 to 1000.

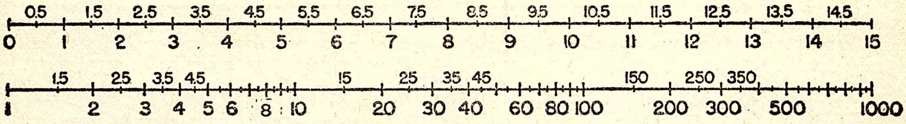


Fig. 13

In the upper scale, half-way intervals are marked, so that we see that 1.5 is exactly half-way between 1.0 and 2.0. In the lower scale, the spacings are not identical, but follow logarithmic principles. Here 1.5 is considerably nearer to 2 than it is to 1; 15, in the middle portion, is the same distance from 10 that 1.5 is from 1; similarly, 150, in the third portion, is at a corresponding distance from 100. In using logarithmic scales, we may let the first portion of our scale represent any multiple of 1 to 10 that we desire; succeeding portions will represent, in order, the next higher multiples of 10.

The facts of Table 12 plotted on a logarithmic scale are shown in Fig. 14, where the paper is placed in such a way that the logarithmic spacings, representing the ratios, are on the X-axis (horizontal) and the standard spacings, representing frequencies, are on the Y-axis (perpendicular). Indicating the ratios on the X-scale, we place the frequencies on the Y-scale just as we did in the coördinate graph of Fig. 12, but the slopes of the lines are somewhat different, so that the picture will be interpreted in a slightly different but more accurate fashion.

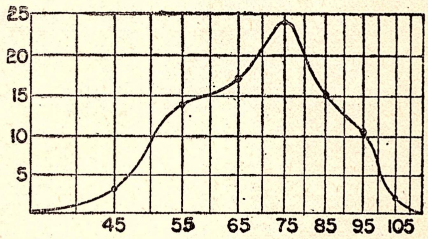


Fig. 14

While we are on the subject of logarithmic graphs, we shall digress for a moment to take note of a case which calls for using the logarithmic scale in the upright position (the Y-axis). Figs. 15 and 16 are representations of the populations of New York City and Chicago, respectively, at 10-year intervals from 1870 to 1940. In Fig. 15, the figures are plotted on a coördinates graph, while in Fig. 16 the same facts are depicted on a semi-logarithmic graph. Since we are chiefly concerned, in comparing population

growth, in the *rate* of change rather than the *amount* of change, the slopes shown in Fig. 16, in which the rates are held to a constant proportion by use of the logarithmic ruling, assist the eye to make a clearer interpretation.

The harmonic mean

For averaging time rates, the harmonic mean gives us some advantages. This mean is found by taking the reciprocal (p. 167) of the arithmetic mean of the reciprocals of the values with which we are dealing. In words, this sounds intricate; in formula, it becomes clearer:

$$\frac{1}{M_H}=\frac{\Sigma\left(\frac{1}{r_1}+\frac{1}{r_2}+\frac{1}{r_3}+\ldots+\frac{1}{r_n}\right)}{n}. \qquad \text{V}$$

Let us suppose that one workman is found to require 10 minutes to

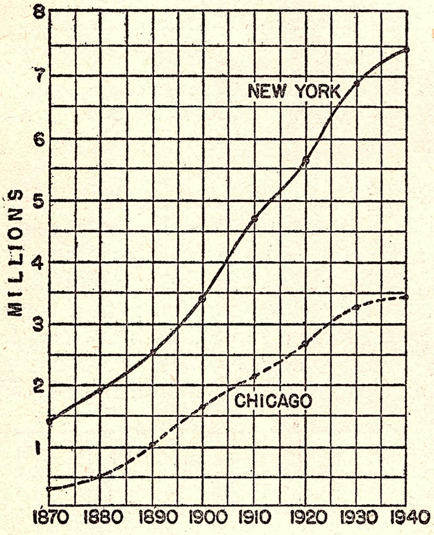


Fig. 15

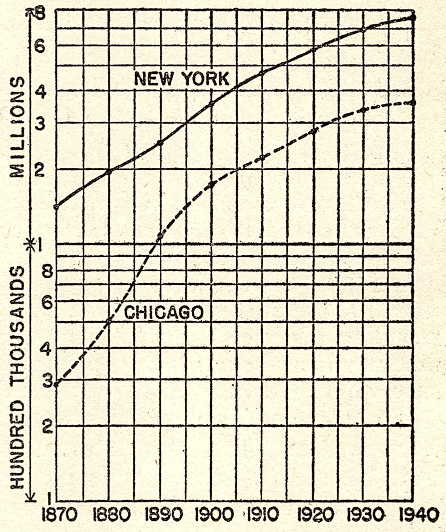


Fig. 16

turn out a piece of work, another requires 12 minutes, another 15 minutes, and the fourth requires 20 minutes.

By obtaining the arithmetic mean, we find that

$$\frac{10+12+15+20}{4}=\frac{57}{4}=14.25 \text{ minutes}$$

represents the "average" time of the group to produce a unit.

Considering the facts differently, we find the output per hour to be:

	Units per Hour
First man	6
Second man	5
Third man	4
Fourth man	3
Total	18

Dividing 18 by 4, the number of men, we get 4.5 units per man as the "average" output per hour. Then $\frac{60}{4.5}=13.333\ldots$ would seem to be

the "average" time per unit. We have here a discrepancy of nearly a whole minute per unit, enough to make a serious difference in a week's output. How may we be sure which, if either, of these figures may be relied upon? This is a case where the harmonic mean will help.

$$\frac{1}{M_H}=\frac{\Sigma\left(\frac{1}{10}+\frac{1}{12}+\frac{1}{15}+\frac{1}{20}\right)}{4}=\frac{\frac{6}{60}+\frac{5}{60}+\frac{4}{60}+\frac{3}{60}}{4}=\frac{\frac{18}{60}}{4}=\frac{4.5}{60}=0.075.$$

Then $M_H = 13.333\ldots.$

To avoid the necessity of making a fresh computation each time we need to find the reciprocal, as in the last step above, we may make use of a table of reciprocals, such as Table C. The table as presented on page 1256 gives values of the reciprocals for all numbers from 1 to 10 at intervals of 0.1. We may extend its range considerably merely by moving the decimal point.*

Finding that the reciprocal of 7.5 is 0.13333, we may extend our reasoning as follows:

n	$\frac{1}{n}$	n	$\frac{1}{n}$
7.5	0.13333	7.5	0.13333
75.0	0.013333	0.75	1.3333
750.0	0.0013333	0.075	13.333
7500.0	0.00013333	0.0075	133.33
etc.		etc.	

This procedure holds, then, for all decile multiples and sub-multiples of the sequence of numbers presented.

Let us take an example of number sequence which does not appear in the table, such as 3.24. From the table, we find the values for $\frac{1}{3.2}$ and $\frac{1}{3.3}$. It is apparent that the value of $\frac{1}{3.24}$ will lie between these two values. Interpolating after our usual fashion (p. 95), we may achieve a reasonable approximation to the desired value. For extensive work, of course, use of larger and more finely graded tables will eliminate the computation of interpolations.

The quadratic mean

One further "average" which is frequently employed in statistical work is the quadratic mean, found by taking the square root of the sum of the deviations from the arithmetic mean divided by the number of items:

$$M_Q=\sqrt{\frac{\Sigma(d)^2}{n}} \qquad \textbf{VIa}$$

where d = the deviation from the arithmetic mean.

*Remembering that the reciprocal of a given number is the number which, when multiplied by the given number, produces a result of 1, we may see that, when we move the decimal point of n one place to the left, we must move the decimal point of $\frac{1}{n}$ one place to the right, and *vice versa*.

A slightly more usable formula for grouped data, making it unnecessary to select the exact arithmetic mean, is:

$$_aM_Q = \sqrt{\frac{\Sigma f(d_a)^2}{n} - \left(\frac{\Sigma fd_a}{n}\right)^2} \qquad \textbf{VIb}$$

where d_a = the deviation of the mid-point of each class interval from the mid-point of an arbitrarily selected group.

For this reason, the quadratic mean is sometimes known as the standard deviation (designated by the G eek letter, σ). It is in this form that we shall meet it again on page 1210.

The principal facts relating to the various forms of average are tabulated, for convenience of reference, in Fig. 17.

MEASURES OF CENTRAL TENDENCY

Designation	Formula	Use	Advantages	Disadvantages
Arithmetic mean	I		Simply computed. Can be employed in further algebraic manipulations.	Influenced by extreme values.
Weighted average			Overcomes distortion to some extent.	Still shows influence of atypical values.
Median	II	Determining the typical "average" case. When original observations are symmetrically distributed, $M_A = M_d = M_O$. When distribution has more low values, $M_A > M_d > M_O$. When distribution has more high values, $M_A < M_d < M_O$.	Simply calculated. More typical than the arithmetic mean. Influenced by the number rather than the size of extreme variations.	Not generally familiar. Cannot be used in further computations. Larger standard and probable errors. (See pp. 1234-1236.)
Mode	III		Most typical and most descriptive. Can be approximated by observation. Less affected than the median by extreme items.	Difficult to ascertain exactly. Not capable of further manipulation. Significance limited when number of cases is small.
Geometric mean	IV	Computing rates of change or index numbers.	Gives consistent results when used to average ratios. Less affected than the arithmetic mean by extreme values.	Relatively difficult to compute. Cannot be determined when zero or negative values appear in the series.
Harmonic mean	V	Averaging rates when the variable factor should be made constant.	Holds constant a factor which appears variable in the data.	Cannot be determined when zero appears in the series.
Quadratic mean	VI	Computing standard deviation.		

Fig. 17

TEST YOUR ABILITY TO COMPUTE OTHER MEASURES OF CENTRAL TENDENCY

9 Compute the geometric means for the figures in Tables α and β. Compare these with your previous computation of the arithmetic means in answer to problem 5.

10 Compute the geometric mean for the data in Table γ, which shows the net earnings of Federal Reserve Banks each year from 1926 to 1941. Plot the figures in Table γ on (a) coördinates graph paper; (b) semilogarithmic graph paper.

11 Table δ shows the number of minutes required by each workman in a plant to complete a finished unit of work. Compute the harmonic mean.

12 Workers in a mailing room completed the following insertions per hour: 771, 946, 612, 1025, 997, 869, 743, 894, 659, 791, 1009, 880, 828, 974, 1132, 811, 907, 783, 842. Using your knowledge of measures of central tendency, prepare a report showing the average expected performance per day and the estimated total output of the mailing room in a 40-hour week. Note the arithmetic differences between the harmonic mean, the arithmetic mean, the median, and the mode, and note also the amount of variance to which these differences will expand when applied to the weekly output.

TABLE γ

Year	Earnings
1926	$16,612
1927	13,048
1928	32,122
1929	36,403
1930	7,988
1931	2,972
1932	22,314
1933	7,957
1934	15,231
1935	9,437
1936	8,512
1937	10,801
1938	9,582
1939	12,243

TABLE δ

Minutes per Unit	Frequency
7	3
8	7
9	15
10	11
11	8
12	2

MEASURES OF DISPERSION

Even though we have arrived at a typical value, one or another of the measures of central tendency discussed in the preceding section, we shall find that it is of little use to us unless we know something about the distribution of cases. An "average" of widely scattered figures may mean nothing at all. In a case where several very large or very small items tend to distort the average, it may be desirable to eliminate these before taking an adjusted average.

In many businesses, a few executives are paid much higher salaries than are any of the other employees. To average in the salaries of these few individuals would raise the computed average of the entire group to a level which would not be representative of the group as a whole.

As a case in point, let us take the actual salary schedule of a small school system, in which the teaching force was paid as follows:

Superintendent	$4000	Elementary-school principal	$1250
High-school principal	3000	Elementary-school principal	1250
Athletic coach-teacher	2400	Elementary-school principal	1250
High-school teacher	1800	Elementary-school teacher	1200
High-school teacher	1800	Elementary-school teacher	1200
High-school teacher	1800	Elementary-school teacher	1200
High-school teacher	1600	Elementary-school teacher	1200
High-school teacher	1600	Elementary-school teacher	1200
High-school teacher	1300	Elementary-school teacher	1100
High-school teacher	1300	Elementary-school teacher	1000

When a question of raising teachers' salaries came before the finance committee of the employing town, the chairman of that committee argued that the teaching force was reasonably well paid, as compared with other towns in that vicinity. Taking the total salary pay-roll of $32,450 and dividing by the total of teaching employees, 20, he arrived at $1622.50 as the "average" salary paid the teachers in the community.

The teachers, on the other hand, contended that there was a big gap between the salaries paid the full-time classroom teachers and the salaries paid the men in the three top-salaried positions, which involved administrative as well as teaching duties. It would be fairer to exclude these three figures as atypical before proceeding to determine the "average" salary paid the teachers. The 17 whose duties were chiefly teaching (including the elementary-school principals, who in this system were also classroom teachers) received a total annual salary of $23,050, an average of $1355.88. The figures were further broken down to show that the 7 high-school teachers received a total of $11,200, an average of $1600.00, while the 10 elementary-school teachers received a total of $11,850, an average of $1185.00. This is a good example of the way in which the inclusion of figures not strictly comparable may lead to conclusions which, even if mathematically correct, have little validity.

The fact that the average age of a man aged 39 and his son aged 3 was 21 would certainly not be interpreted as giving both of them the right to register as voters!

The range

Before beginning to compute figures, then, it is obvious that we must first subject our items to careful scrutiny, to determine to what extent they will help us to arrive at a reliable guide. First of all, we may consider the spread of the figures.

The range is simply the spread between the largest and the smallest figures. It is expressed either as a span from the minimum to the maximum or as the result of subtracting the minimum from the maximum.

$$R = a_{min}\text{-}a_{max} \qquad \textbf{VIIa}$$

$$R = a_{max} - a_{min} \qquad \textbf{VIIb}$$

As the maximum and the minimum are easily discernible by a mere inspection of the figures, the range is obtained very easily. Since no other figures are taken into account, however, the range may be affected tremendously by an "unusual" case at either extremity of the distribution (as in the first calculation of teachers' salaries).

For the case presented in Table 2, we find that the range is

$$R = \$1.11\text{-}\$3.89$$

or

$$R = \$3.89 - \$1.11 = \$2.88.$$

For the case considered in Table 7, we find a much smaller range,

$$R = 464\text{-}472$$

or

$$R = 472 - 464 = 8.$$

Again, in the case shown in Table 10, the range is

$$R_a = 22.5\text{-}67.5$$

or

$$R_a = 67.5 - 22.5 = 45.$$

Note that, in this last, case, we take the extremes of the class intervals, even though we are not certain, from the facts at hand, that any case actually drops as low as 22.5 or rises as high as 67.5. This, then, is only an approximate range and is so indicated by the subscript, *a*, standing for "approximate".

In the same way, for Table 11, we arrive at

$$R_a = 1150\text{-}2150$$

or

$$R_a = 2150 - 1150 = 1000.$$

Quartile deviation

To avoid some of the effects of extreme values on the range, we may divide our figures into four equal groups, marked by quartiles. The first quartile (designated as Q_1) is located at the point where 25 per cent of the items fall below and 75 per cent lie above; Q_2 is the point at which 50 per cent of the items lie below and 50 per cent rise above; Q_3 is the point at which 75 per cent of the items appear below and 25 per cent above.

Since $Q_2 = M_d$ in a symmetrical distribution, we may simply take the formula for the median (IIa) and repeat it as our formula for the second quartile:

$$Q_2 = \frac{n}{2}\,. \qquad \textbf{VIII}$$

Since Q_2 represents 50 per cent of the cases and Q_1 represents 25 per cent (or one-half that many), it is obvious that

$$Q_1 = \frac{n}{4}\,. \qquad \textbf{IX}$$

Then the third quartile may be found by multiplying the first quartile by 3, so that

$$Q_3 = \frac{3\ (n)}{4}\,. \qquad \textbf{X}$$

By definition, it is obvious that 50 per cent of the values lie between the first and the third quartiles. The difference between the two quartiles, called the interquartile range, is expressed by

$$\text{IQR} = Q_3 - Q_1\,. \qquad \textbf{XI}$$

A small interquartile range indicates a large degree of uniformity, while a large interquartile range indicates a lesser degree of uniformity.

The semi-interquartile range is, as its name indicates, merely the IQR divided by 2. This value, known as the *quartile deviation*, affords

a means of measuring the average distance of each quartile from the median. To secure this value, we divide the difference between Q_3 and Q_1 by 2.

$$\mathrm{QDv}=\frac{Q_3-Q_1}{2}. \qquad \text{XII}$$

Measuring this value along the line of range in either direction from the median gives us a range which will include roughly fifty per cent of the cases. In a symmetrical distribution, the median plus or minus the quartile deviation will include exactly fifty per cent of the cases.

In comparing the dispersions in two separate distributions, we may make use of either the coefficient of variation or the coefficient of quartile deviation, the latter known also as the coefficient of dispersion. These are relatively simple to use and their computation follows logically on the formulas which we have just been considering:

$$_c\mathrm{D}_{\mathrm{sp}}=\frac{Q_3-Q_1}{Q_3+Q_1} \qquad \text{XIII}$$

$$_cV=\frac{\sigma}{M_A}\cdot 100. \qquad \text{XIV}$$

By the use of either of these tests for any measure of relative variation, we may determine how closely the two or more distributions resemble each other, even though they are expressed in different units.

TABLE 13

SALES-TICKET QUARTILES AND DECILES

	Price		Occurrences	
	1.11		1	
	1.16		1	
	1.25		2	
	1.27		1	
	1.40		1	
	1.48		1	
	1.51		2	
	1.55		1	
	1.61		1	
	1.73		1	
	1.75		3	D_1
	1.80		2	
	1.81		2	
	1.83		3	
	1.89		3	
	1.95		3	
	2.01		3	D_2
	2.11		3	
	2.13	Q_1	4	
	2.21		4	
	2.25		4	D_3
	2.28		5	
	2.31		7	D_4
	2.45		8	
M_d	2.50	Q_2	12	D_5
	2.55		8	
	2.69		7	D_6
	2.72		5	
	2.75		4	D_7
	2.79		4	
	2.87	Q_3	4	
	2.89		3	
	2.99		3	D_8
	3.05		3	
	3.11		3	
	3.17		3	
	3.19		2	
	3.20		2	
	3.25		3	D_9
	3.27		1	
	3.39		1	
	3.45		1	
	3.49		2	
	3.52		1	
	3.60		1	
	3.73		1	
	3.75		2	
	3.84		1	
	3.89		1	

Decile range

In determining the decile range, we divide the values into ten groups, each containing one-tenth of the frequencies. The computation is similar to the formulas for computing quartiles:

$$D_1=\frac{n}{10} \qquad \text{XV}$$

and so on up to

$$D_9=\frac{9n}{10}. \qquad \text{XVI}$$

Note here the special case, D_5, in a symmetrical distribution:

$$D_5=\frac{5n}{10}=\frac{n}{2}=Q_2=M_d. \qquad \text{XVII}$$

The decile 1-9 range, also called the 10-90 percentile range (since a decile divided by 10 equals a percentile and hence a decile is 10 times a percentile), indicates the points between which 80 per cent of the values lie.

$$\text{PCR} = D_9 - D_1 . \qquad \textbf{XVIII}$$

Subjecting the values in Table 3 to interpretation by formulas XVIII and XI, we get the results shown in Table 13.

Note here that we have merely approximated the location of the decile and the quartile points. Where the number of occurrences up to the desired point indicates that some one of the three items is the first decile (the price, $1.75, occurs 3 times, it will be noted), we might interpolate to determine how much of the difference between $1.75 and the next item represented the exact mathematical location of the first decile. Here the variations in price are extremely small as between one item and the next and the number of occurrences is not great, so that this approximation will not materially distort our final figures.

By substitution in the formulas, we have:

$$\text{PCR} = D_9 - D_1 = 3.25 - 1.75 = 1.50$$

$$\text{IQR} = Q_3 - Q_1 = 2.87 - 2.13 = 0.64.$$

For the sake of comparison, let us set down the facts regarding the range, the interquartile range, and the 10-90 percentile range in parallel columns:

	Whole Range	10-90 Percentile Range	Interquartile Range
Formula employed	VIIb	XVIII	XI
Value	2.78	1.50	0.64
Per cent of cases included	100	80	50

Thus, we see that one-half of our cases fall within a relatively small range (less than one-quarter of the whole range) and that 80 per cent of the cases fall within a greatly reduced range (very little over one-half of the whole range). By shearing off the items in Table 13 which appear before we reach D_1 and after we have reached D_9, we have cut the length of the table materially yet we have not disposed of a very great number of cases. If we consider on y that portion of Table 13 which l es between Q_1 and Q_3, we have very greatly compressed the table.

Interpolation to get the exact locations of Q_1, D_1, etc. would affect these figures slightly, but the amount of increase in accuracy would scarcely warrant the time spent in additional computation, as the reader who practices on this illustration will readily discover for himself.

Mean deviation

In order to determine the amount of deviation involved, we may compute for the whole number of items the average distance from a given measure of central tendency. Obviously, since there are approximately the same number of these on either side of the central measure, their sum would be 0 or close to 0 if we took the signs into account. For this purpose, then, we decide to ignore the signs, expressing both $+26$ and -26 by the symbol, $|26|$, which is an abbreviated way of indicating the *pure value* as distinguished from the *signed value* of the numeral.

$$Dv_{M_A}=\frac{\Sigma|x|}{n} \qquad \textbf{XIXa}$$

where $|x|$ = the deviation of a value from the arithmetic mean

and

$$Dv_{M_d}=\frac{\Sigma|d|}{n} \qquad \textbf{XIXb}$$

where $|d|$ = the deviation of a value from the median.

For a fairly simple case, let us consider the record of completed units of work in a factory department, as given in the first two columns of Table 14.

TABLE 14

WORK PERFORMANCE OF 9 EMPLOYEES

WORKMAN	UNITS COMPLETED m	DEVIATIONS FROM ARITHMETIC MEAN $\|x\|$	DEVIATIONS FROM MEDIAN $\|d\|$
A	41	41.00−36.44=4.56	41−37=4
B	29	36.44−29.00=7.44	37−29=8
C	37	37.00−36.44=0.56	37−37=0
D	44	44.00−36.44=7.56	44−37=7
E	31	36.44−31.00=5.44	37−31=6
F	35	36.44−35.00=1.44	37−35=2
G	40	40.00−36.44=3.56	40−37=3
H	32	36.44−32.00=4.44	37−32=5
I	39	39.00−36.44=2.56	39−37=2
$n=9$	$\Sigma m=328$	$\Sigma\|x\|=37.56$	$\Sigma\|d\|=37$

To find the arithmetic mean, we revert to formula Ia, finding

$$M_A=\frac{\Sigma m}{n}=\frac{328}{9}=36.44.$$

We then compute the items for column 3 by subtracting each item in column 2, in turn, from the arithmetic mean if the item is smaller than the mean, or by subtracting the mean from it if the item is larger than the mean. Adding the items, we get a total of 37.56.

Then

$$Dv_{M_A}=\frac{\Sigma|x|}{n}=\frac{37.56}{9}=4.17.$$

To find the median, we use formula IIa, finding

$$M_d=\frac{n+1}{2}\text{th}=\frac{9+1}{2}\text{th}=\frac{10}{2}\text{th}=5\text{th}\ .$$

Arranging the cases in order, we have:

29, 31, 32, 35, 37, 39, 40, 41, 44.

The fifth case, counting from either end, gives us 37 as the median. We may now compute the items for column 4 by comparing the items in

column 2 with the median, subtracting the larger number from the smaller in each case. Adding these deviations, we have a total of 37.

Then $$Dv_{M_d}=\frac{\Sigma|d|}{n}=\frac{37}{9}=4.11.$$

It should be noted that the average deviation from the median will always be smaller than the average deviation from any other measure of central tendency.

When we are dealing with data grouped in class intervals, we use the mid-points of the class intervals as the points from which to compute the deviations. Then we multiply each computed deviation by the frequency of its class interval and divide the total of the product by the number of cases.

$$_fDv_{M_A}=\frac{\Sigma(f|m_x|)}{n} \qquad \text{XIXc}$$

and $$_fDv_{M_d}=\frac{\Sigma(f|m_d|)}{n} \qquad \text{XIXd}$$

Treating the case presented in Table 11 by formula XIX, we find the results shown in Table 15.

TABLE 15

ANNUAL SALARIES OF EMPLOYEES

Class Interval	Mid-point	Frequency	Deviations from Arithmetic Mean	Product of Columns 3 and 4
i	m	f	$\mid m_x \mid$	$f \mid m_x \mid$
$1150-1250	$1200	4372	465	2032980
1250-1350	1300	3299	365	1204135
1350-1450	1400	3003	265	795795
1450-1550	1500	3587	165	591855
1550-1650	1600	4092	65	265980
1650-1750	1700	11129	35	389515
1750-1850	1800	18578	135	2508030
1850-1950	1900	3238	235	760930
1950-2050	2000	1971	335	660285
2050-2150	2100	1303	435	566805
$n=10$		$\Sigma f=54572$	$\Sigma \mid m_x \mid =2500$	9776310

Here the arithmetic mean had been found (p. 1196) to be 1664.91, which we may consider for all practical purposes to be 1665. To fill column 4, subtract from 1665 each mid-point up to and including 1600 and then continue by subtracting 1665 from each mid-point from 1700 on to the end of the list. Since our intervals are uniformly 100, notice that the sum of the deviations, beginning with the two in the middle and making up pairs progressively from the middle, will be in multiples of 100:

$$35+65=100$$
$$135+165=300$$
etc.

This observation considerably shortens the task of computing $\Sigma \mid m_x \mid$.

We next multiply these items by the frequencies of their various classes, setting down the products in column 5 and totalling.

Then $$_fDv_{M_A}=\frac{\Sigma f|m_x|}{\Sigma f}=\frac{9776310}{54572}=179.$$

The method for formula XIXd is similar. Its solution is left to the reader.

Use of formulas XIXc and XIXd involves rather laborious computations. Since the class intervals are uniform in size, we may simplify the calculations by performing the computations in terms of class intervals and then multiplying the results by the size of the class interval. As the method is similar to that of the work just shown, we leave the actual computation to the reader.

Standard deviation

More used than the mean deviation, the standard deviation gives more emphasis to the extreme variations because it is based on the squares of the deviations and, since all squares are positive, the extreme values thus take on added importance. The symbol used to express standard deviation is the lower-case Greek letter, sigma, σ, and deviations are usually expressed in terms of "one sigma", "two sigmas", etc.

Since the standard deviation is expressed by taking the quadratic mean of the deviations from the arithmetic mean, we are not surprised to find that its basic formula resembles that for the quadratic mean (cf. formula VIa):

$$\sigma=\sqrt{\frac{\Sigma(d_m)^2}{n}}. \qquad \text{XX}$$

When dealing with a frequency distribution arranged by class intervals, we assume as a mean the mid-point of the class near the peak or concentration point of the data. Then we record the deviations of all other mid-points from the assumed mean, counting the number of class intervals by which each mid-point is removed from it.

We next multiply the signed value of each deviation by its class frequency, take the total of the fd_m values, and divide by the total of frequencies. This gives the correction factor, c. (If we have happened to select the exact mean as the assumed mean, the value of c will be 0.)

$$c=\frac{\Sigma fd_m}{\Sigma f} \qquad \text{XXI}$$

After squaring the deviations and adding the results, we divide by the total number of frequencies. From this number, we subtract the square of the correction factor and then take the square root of the remainder. This square root, multiplied by the size of the class interval, gives us our σ value:

$$\sigma=i\sqrt{\frac{\Sigma f(d_m)^2}{\Sigma f}-c^2}. \qquad \text{XXII}$$

Table 5 treated in this fashion produces the results shown in Table 16.

Since in this case $\Sigma fd_m=0$, the correction factor, $c=\frac{\Sigma fd_m}{\Sigma f}$, will be $c=\frac{0}{144}=0$ and $c^2=0$.

Now substituting in formula XXII, we have

$$i=0.25 \quad \text{and} \quad \Sigma f=144$$

so that

$$\sigma=i\sqrt{\frac{\Sigma f(d_m)^2}{\Sigma f}-c^2}=0.25\sqrt{\frac{806}{144}-0}=0.25\sqrt{5.6}=0.25(\pm 2.04)=\pm 0.51.$$

Then our arithmetic mean, previously determined as 2.50 (p. 1187), plus or minus the standard deviation, 0.51, should give us the range within which we expect 68.27 per cent of the cases to fall.

$$2.50\pm 0.51=1.99 \text{ or } 3.01.$$

TABLE 16

SALES-TICKET DEVIATIONS

CLASS INTERVAL	MID-VALUE	FRE-QUENCY	DEVIATION FROM ASSUMED MEAN	SQUARE OF DEVIATION	PRODUCT OF COLUMNS 3 AND 4	PRODUCT OF COLUMNS 3 AND 5
i	m	f	d_m	$(d_m)^2$	fd_m	$f(d_m)^2$
\$0.875-1.125	1.00	1	−6	36	− 6	36
1.125-1.375	1.25	4	−5	25	−20	100
1.375-1.625	1.50	6	−4	16	−24	96
1.625-1.875	1.75	11	−3	9	−33	99
1.875-2.125	2.00	12	−2	4	−24	48
2.125-2.375	2.25	24	−1	1	−24	24
2.375-2.625	2.50	28	0	0	0	0
2.625-2.875	2.75	24	1	1	24	24
2.875-3.125	3.00	12	2	4	24	48
3.125-3.375	3.25	11	3	9	33	99
3.375-3.625	3.50	6	4	16	24	96
3.625-3.875	3.75	4	5	25	20	100
3.875-4.125	4.00	1	6	36	6	36
		$\Sigma f=144$			$\Sigma fd_m=0$	$\Sigma f(d_m)^2=806$

Taking the slightly less symmetrical distribution of Table 12, we have the results shown in Table 17.

TABLE 17

COMPANY RATIOS—DEVIATIONS

CLASS INTERVAL	MID-POINT	FRE-QUENCY	DEVIATION FROM ASSUMED MEAN	SQUARE OF DEVIATION	PRODUCT OF COLUMNS 3 AND 4	PRODUCT OF COLUMNS 3 AND 5
i	m	f	d_m	$(d_m)^2$	fd_m	$f(d_m)^2$
40- 50	45	3	−3	9	− 9	27
50- 60	55	14	−2	4	−28	56
60- 70	65	17	−1	1	−17	17
70- 80	75	24	0	0	0	0
80- 90	85	15	1	1	15	15
90-100	95	11	2	4	22	44
100-110	105	2	3	9	6	18
		$\Sigma f=86$		$\Sigma(d_m)^2=28$	$\Sigma fd_m=11$	$\Sigma f(d_m)^2=177$

Before we make use of these figures, we may obtain a simple check (the Charlier check) on their accuracy by taking one class higher as our assumed mean. Expressed mathematically, this point would be ${}_iM_A={}_iM_A+1$. Deviations computed from this second point would be indicated by d_m+1. It should at once be evident that

$$\Sigma f(d_m+1)^2=\Sigma f(d_m)^2+2\Sigma fd_m+\Sigma f. \qquad \textbf{XXIII}$$

By making the computations in Table 18, we may verify our figure for $\Sigma f(d_m)^2$ before entering into further computations with it.

Formula XXIII tells us that the value for $\Sigma f(d_m+1)^2$, 241 as found in Table 18, should be equal to the sum of $\Sigma f(d_m)^2$ and Σf increased

by two times $\Sigma f d_m$. All of these values may be found from Table 17 and inserted in the formula, getting

$$241 \stackrel{?}{=} 177+86+2(-11)=263-22=241.$$

TABLE 18

COMPANY RATIOS—CHARLIER CHECK

CLASS INTERVAL i	FRE-QUENCY f	DEVIATION FROM AS-SUMED MEAN d_m	DEVIATION FROM NEXT HIGHER CLASS d_m+1	SQUARE OF DEVIATION $(d_m+1)^2$	PRODUCT OF COLUMNS 2 AND 5 $f(d_m+1)^2$
40- 50	3	−3	−2	4	12
50- 60	14	−2	−1	1	14
60- 70	17	−1	0	0	0
70- 80	24	0	1	1	24
80- 90	15	1	2	4	60
90-100	11	2	3	9	99
100-110	2	3	4	16	32
					$\Sigma f(d_m+1)^2=241$

Returning now to formula XXII, we substitute this value for $\Sigma f(d_m)^2$, as computed in Table 17 and verified by formula XXIII and the computations of Table 18, and the value for c^2, as previously found by formula XXI:

$$\sigma=i\sqrt{\frac{\Sigma f(d_m)^2}{\Sigma f}-c^2}=10\sqrt{\frac{177}{86}-\left(-\frac{11}{86}\right)^2}=10\sqrt{\frac{177}{86}-\frac{121}{7396}}$$

$$=10\sqrt{\frac{15222-121}{7396}}=10\sqrt{\frac{15101}{7396}}=10\sqrt{2.055}=10\pm1.433=\pm14.33.$$

We find, then, that 14.33 is the amount which should be read on either side of the mean in order to get 68.27 per cent of the cases,

$$75\pm14.33=60.67 \text{ or } 89.33.$$

All numbers in the table between 60.67 and 89.33 are within a deviation of one sigma from the mean

The relative merits of the va ious measures of dispersion are summarized in Fig. 18 (p. 1213).

TEST YOUR KNOWLEDGE OF MEASURES OF ABSOLUTE DISPERSION

Using Tables α and β, determine for each:

13 The range

14 The quartile deviation

15 The 10-90 percentile

16 The mean deviation

17 The standard deviation

Skewness

As we have already noted, not all frequency distributions have their peaks, or greatest number of frequencies, at the exact center of the diagram. In some cases, the peak may be to the left (indicating a greater number of low values), the curve tapering off toward the right (the high values); in other cases, the peak may be toward the right, with the curve tapering off toward the left. This tendency is expressed as skewness, since the curve skews in one direction or the other. The curves shown in Figs. 19 and 20, based on Tables 19 and 20, are examples of skewed curves.

If we desire to compare two or more distributions, we may do so safely if we know to what extent the distributions are comparable, even

MEASURES OF ABSOLUTE DISPERSION

Designation	Formula	Cases Included	Advantages	Disadvantages
Over-all range	VII	100%	Simple obtained. Easily understood.	Least satisfactory measure.
Quartile deviation	XII	50%†	Concentrates attention on central cases.	Cannot be used in further computation. Not a measure of central tendency for asymmetrical distributions.
10-90 percentile	XVIII	80%	Avoids defects of the over-all range by excluding extremes.	Cannot be used in further computation.
Mean deviation*	XIX	57.5%‡	Depends on every value in series, but is less influenced by extremes. May be computed about any measure of central tendency (sum of deviations from median is least). Ignores signs in summing deviations. May be used in further algebraic computations.	Of slight utility when variation of values is large.
Standard deviation*	XXII	$\pm 1\sigma$: 68.27% $\pm 2\sigma$: 95.45% $\pm 3\sigma$: 99.73%	Subject to further algebraic computation. Affected by value of every item. Deviations from arithmetic mean yield minimum sum.	Gives greater emphasis to extreme values because deviations are squared.

* In normal distribution.

† If laid off in each direction from $\frac{Q_1+Q_2}{2}$.

‡ Limits obtained for $M_A \pm \text{Dv}M_A$

Fig. 18

though they may be based on different class intervals or ranges. As a working rule, we may fairly safely take the statement that the mode is equal to the arithmetic mean decreased by three times the difference between the arithmetic mean and the median.

$$M_o = M_A - 3(M_A - M_d). \qquad \text{XXIV}$$

Illustrative Example

For purposes of understanding skewness, we are taking a hypothetical distribution of the salaries of clerical workers in an office, as shown in Table 19.

Plotting the m values on the X-axis and the f values on the Y-axis, we get the curve shown in Fig. 19. Applying formulas Id, IIb (with IIc for check), and III, we get:

$$M_A = \frac{3285}{132} = 24.89 \qquad M_d = 22.50 + \frac{66-41}{49} \cdot 5 = 25.05$$

$$M_o = 22.50 + \frac{32}{32+24} \cdot 5 = 25.36 \qquad M_d = 27.50 - \frac{66-42}{49} \cdot 5 = 25.05$$

If, for the sake of argument, we take a different arrangement of the salary distribution for these same workers, in which we hand out some

TABLE 19

SALARIES OF CLERICAL WORKERS

CLASS INTERVAL	MID-POINT	FRE-QUENCY	PRODUCT OF COLUMNS 2 AND 3	CUMULATED FREQUENCIES	
i	m	f	fm	<	>
\$12.50-17.50	15	17	255	17	132
17.50-22.50	20	24	480	41	115
22.50-27.50	25	49	1225	90	91
27.50-32.50	30	32	960	122	42
32.50-37.50	35	7	245	129	10
37.50-42.50	40	3	120	132	3
		$\Sigma f=132$	$\Sigma fm=3285$		

salary increases in such a way that our list of frequencies is exactly reversed, we note that we have a curve in which the area covered is

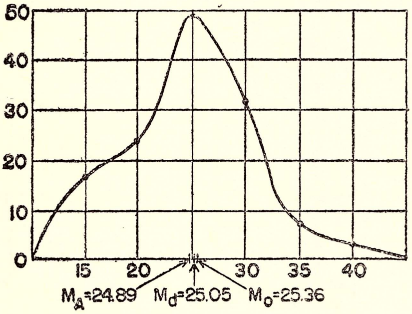


Fig. 19

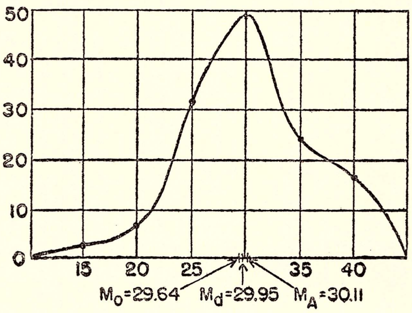


Fig. 20

exactly the same, but that the curve is the exact reverse of the one plotted in Fig. 19. This curve (Fig. 20) is based on Table 20.

TABLE 20

SALARIES OF CLERICAL WORKERS

CLASS INTERVALS	MID-POINT	FRE-QUENCY	PRODUCT OF COLUMNS 2 AND 3	CUMULATED FREQUENCIES	
i	m	f	fm	<	>
\$12.50-17.50	15	3	45	3	132
17.50-22.50	20	7	140	10	129
22.50-27.50	25	32	800	42	122
27.50-32.50	30	49	1470	91	90
32.50-37.50	35	24	840	115	41
37.50-42.50	40	17	680	132	17
		$\Sigma f=132$	$\Sigma fm=3975$		

Using the same formulas as in the case of Table 19, we arrive at:

$$M_A=\frac{3975}{132}=30.11 \qquad M_d=27.50+\frac{66-42}{49}\cdot 5=29.95$$

$$M_o=27.50+\frac{24}{24+32}\cdot 5=29.64 \qquad M_d=32.50-\frac{66-41}{49}\cdot 5=29.95$$

Computing M_o by means of formula XXIV, we get, for Table 19,

$$M_o=24.89-3(24.89-25.05)=24.89-3(-0.16)=24.89+0.48=25.37$$

and, in the case of Table 20,

$$M_o=30.11-3(30.11-29.95)=30.11-3(0.16)=30.11-0.48=29.63,$$

in both cases coming out only one cent different from what we got when

we applied formula III. If you consider greater accuracy desirable in this case, you may retain the fractional values and substitute them in the last step taken.

From this computation, we note that the value of $(M_A - M_d)$ is negative when the curve skews to the left and positive when it skews to the right. Compare this finding with our previous conclusions as to the relationships among the various measures of central tendency (p. 1197).

We may express the skewness as a pure number by dividing the difference between the arithmetic mean and the mode by the standard deviation:

$$\mathrm{Sk} = \frac{M_A - M_d}{\sigma}. \qquad \textbf{XXV}$$

A positive result, following the line of reasoning above, indicates a skew toward high values, while a negative result indicates a skew toward low values. The coefficients of skewness of two distributions must be rather closely in agreement if comparison is to be made between them. If there is a great divergence between the coefficients, this fact should be taken into consideration when drawing conclusions from any comparisons which may be made.

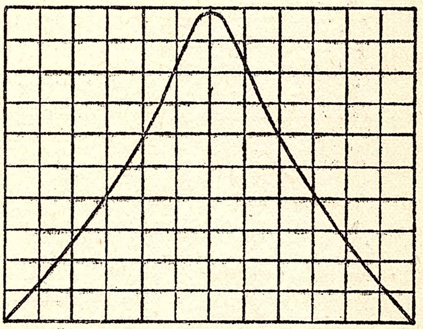


Fig. 21

Kurtosis

Another point to notice in considering the distribution curve is the amount of peakedness or *kurtosis*. The curve of a normal distribution was shown in Fig. 5. When the peak is more pronounced than this, as in Fig. 21, the curve is called *leptokurtic* (sharp or pointed); when the peak is less pronounced, as in Fig. 22, it is called *platykurtic* (flat).

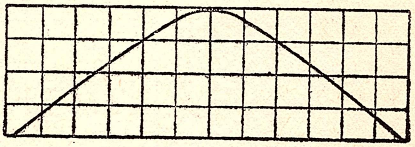


Fig. 22

The degree by which the curve departs from the normal should be kept in mind in drawing inferences. A leptokurtic curve indicates a preponderance of frequencies at a point in the scale; a platykurtic curve indicates little variation amongst a number of class intervals, as shown in Figs. 21 and 22.

In any graphing of curves, however, care must be taken to use scales which sufficiently distinguish the relationships between the values. If the class intervals are brought too close together while the fre-

quencies are stretched out, the result will be a distorted curve such as that shown in Fig. 23. On the other hand, if the class intervals are spaced rather widely, but the frequencies are plotted on lines too close together, the curve tends to be flattened to such an extent that comparisons are not clearly made, as in Fig. 25. Fig. 24 shows the correct drawing of the curve for Table 21, on which all three figures are based. In Fig. 24, the steady and equal rise from class to class is easily interpreted, whereas Fig. 23 seems to indicate a sharp change between classes and Fig. 25 seems to show a barely perceptible increase. The terms, platykurtic and leptokurtic, are not applicable (in the sense just defined) to these mis-drawn curves, but are reserved for discussions of correctly made diagrams.

TABLE 21

MID-POINT m	FREQUENCY f
10	4
20	8
30	12
40	8
50	4

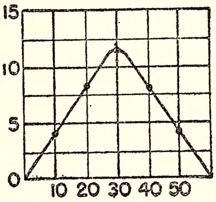


Fig. 23

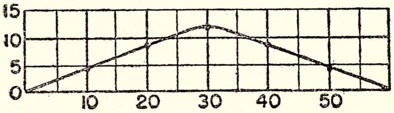


Fig. 24

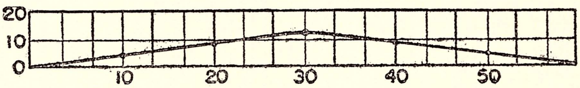


Fig. 25

TEST YOUR KNOWLEDGE OF MEASURES OF RELATIVE DISPERSION

18 Using the values of the arithmetic mean and the median as determined in answer to problems 5 and 6, compute the modes for the data in Tables α and β. Compare your answers with the values for the mode previously found in your answers to problems 7 and 8.

19 Find the coefficients of skewness for the data in Tables α and β.

TIME SERIES AND VARIATION

When we are considering some types of information, in which dates of occurrence or spans of time play an important part, it seems logical to use the names of the months or the years for the designations along the X-axis, plotting the observations on the Y-axis.

In analyzing time series, we are concerned with four distinct movements or changes in the given figures:

a *Seasonal variation*—an upward or downward shift of figures month by month (or day by day), occasioned by the changing seasons and recurring year after year.

b *Secular trend*—growth or decline over a period of years (10 or more).

c *Cyclical movement*—a long climb or decline over a period of several years. (Not all cycles are of the same length and not all are of the same degree of intensity, but frequently a rhythm may be observed at approximately equal intervals of time.)

d *Random variations*—due to unusual occurrences, such as a period of

intense cold, heat, drought, or rain, or the outbreak of war, or an unexpectedly heavy demand. A department store, for example, might find a tremendous drop in expected sales if a blizzard occurred on a day when it had been prepared to welcome an influx of bargain hunters. Because of inability to get a supply of goods, owing to a strike or a bad fire in a warehouse, a store or factory might find itself unable to meet a demand and its sales might drop; simultaneously, one of its competitors might pick up additional business from customers who would normally have patronized the handicapped firm. Sometimes a notation would be written along the ordinates line of the graph depicting the day-to-day or month-to-month figures to help in interpreting the unusualness of the figure. More often, a notation would be attached to the graph. In computing averages over a considerable period—especially in comparing similar dates from year to year—one obtains a better view of the general conditions if such an unusual figure is excluded from the computations.

Seasonal variation

Many activities are of a seasonal nature. Department-store sales are apt to be highest at the Christmas and Easter seasons; agricultural production is naturally greatest in the summer and early autumn; most people purchase their new automobiles in the spring and early summer.

Illustrative Example

Figures published by the Bureau of the Census show the number of passenger automobiles reported as sold by factories in the United States for each month over a 10-year period, as detailed in Table 22. (We are deliberately stopping our comparison of these figures with the year, 1940, because the imminence and presence of war activities in subsequent years and the complete stoppage of new production after the outbreak of war clearly indicates the non-comparability of later figures.) The plotting of these figures on a coördinates graph is shown in Fig. 26. In plotting, only the thousands need be taken into consideration. Despite the many "peaks" and "valleys" in the curve, we get an impression of a distinct upward trend in the number of sales, except for the year, 1938.

Close scrutiny of the diagram discloses that these peaks and valleys recur year after year in substantially the same months, although the actual figures may vary greatly from year to year. These seasonal variations interfere with our over-all view of the trend. We may very easily compute the adjustments which would have to be made to compensate for this factor.

In Table 23, we have recorded the percentages which each month's sales are of the total sales for the year concerned. If each month had the same number of days, and if the sales were uniform month by month, it is evident that the expected percentage of sales for each month would be $\frac{1}{12}$ of the total, or 8.375%. By examining our figures, we discover that eight months in 1935

TABLE 22

FACTORY SALES OF PASSENGER AUTOMOBILES IN THE UNITED STATES—BY MONTHS, 1931 TO 1940

	Total	January	February	March	April	May	June	July	August	September	October	November	December
1931	1,973,090	138,317	180,419	231,244	286,917	271,475	210,396	184,173	155,425	109,228	58,415	49,184	97,897
1932	1,135,491	98,803	94,110	99,399	120,937	157,756	160,338	94,705	75,907	64,748	35,107	47,532	86,149
1933	1,573,512	109,833	90,128	97,469	149,755	180,651	207,597	191,265	191,414	157,376	104,870	42,365	50,789
1934	2,177,919	112,754	186,774	279,274	288,355	273,764	261,280	223,094	183,500	125,040	84,003	49,020	111,061
1935	3,252,244	227,554	273,576	359,410	387,158	305,547	294,182	274,344	181,130	56,097	213,310	336,914	343,022
1936	3,669,528	297,692	224,211	342,870	416,431	384,431	375,337	371,922	209,351	90,101	190,242	341,085	425,365
1937	3,915,889	309,494	296,788	403,879	439,980	425,432	411,414	360,400	311,456	118,671	298,662	295,328	244,385
1938	2,000,985	155,505	139,380	174,065	176,078	154,958	136,531	106,841	58,624	65,159	187,494	320,344	326,006
1939	2,866,796	281,465	243,000	299,703	273,409	237,870	246,704	150,738	61,407	161,625	251,819	285,252	373,804
1940	3,692,328	362,897	337,756	352,922	362,139	325,676	286,040	168,769	46,823	224,470	421,214	407,091	396,531
Total	26,257,782	2,094,314	2,066,142	2,640,235	2,901,159	2,718,050	2,589,819	2,126,251	1,475,037	1,172,515	1,845,136	2,174,115	2,455,009
Averages	2,675,778.2	209,431.4	206,614.2	264,023.5	290,115.9	271,805	258,981.9	212,625.1	147,503.7	117,251.5	184,513.6	217,411.5	245,500.9

TABLE 23

FACTORY SALES OF PASSENGER AUTOMOBILES IN THE UNITED STATES—PER CENT OF ANNUAL SALES EACH MONTH, 1931 TO 1940

	Total*	January	February	March	April	May	June	July	August	September	October	November	December
1931	99.5	7.0	9.1	11.7	14.5	13.8	10.7	9.3	7.9	5.5	3.0	2.5	4.5
1932	100.0	8.7	8.3	8.8	10.7	13.9	14.1	8.3	6.7	5.6	3.1	4.2	7.6
1933	100.2	7.0	5.8	6.2	9.5	11.5	13.2	12.2	12.2	10.0	6.7	2.7	3.2
1934	99.4	5.2	8.5	12.8	13.2	12.6	11.5	10.2	8.4	5.7	3.9	2.3	5.1
1935	100.0	7.0	8.4	11.1	11.9	9.4	9.0	8.4	5.6	1.7	6.6	10.4	10.5
1936	99.9	8.1	6.1	9.3	11.3	10.5	10.2	10.1	5.7	2.5	5.2	9.3	11.6
1937	99.4	7.9	7.6	10.3	11.2	10.2	10.5	9.2	8.0	3.0	7.6	7.5	6.3
1938	100.0	7.8	7.0	8.7	8.8	7.7	6.8	5.3	2.9	3.3	9.4	16.0	16.3
1939	99.9	9.8	8.5	10.5	9.5	8.3	8.6	5.3	2.1	5.6	8.8	9.9	13.0
1940	100.2	9.8	9.1	9.6	9.8	8.9	7.7	4.6	1.3	6.1	11.4	11.0	10.7

* These totals are computed by adding the percentages of the several months. Variations from 100% are due to dropped or raised decimals when the results of the divisions are rounded off to one decimal place. This is an excellent means of checking the accuracy of the monthly computations. Unless the total is fairly close to 100%, there is likelihood that some error has occurred.

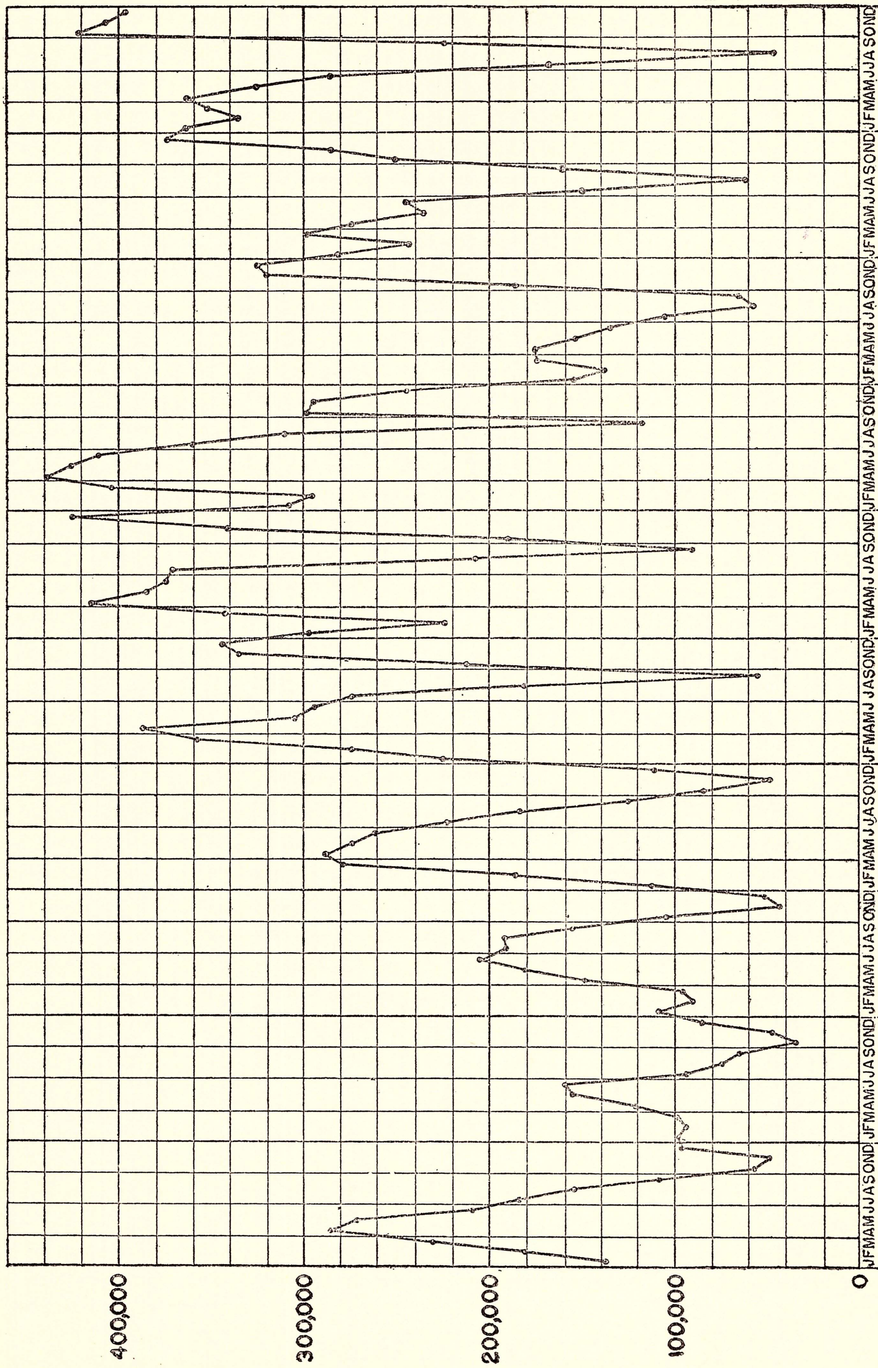

0
100,000
200,000
300,000
400,000
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940

Fig. 26

are above this percentage while the others are below. We may express each monthly percentage as an index by multiplying it by 12, as shown for the 1935 figures in Table 24, column 4. These could be plotted as shown by the solid line in Fig. 27.

TABLE 24

INDEX OF FACTORY SALES OF PASSENGER AUTOMOBILES IN THE UNITED STATES—1935, BY MONTHS

Month	Monthly Sales	Per Cent of Annual Total	Index of Monthly Sales
January	227,554	7.0	84.0
February	273,576	8.4	100.8
March	359,410	11.1	133.2
April	387,158	11.9	142.8
May	305,547	9.4	112.8
June	294,182	9.0	108.0
July	274,344	8.4	100.8
August	181,130	5.6	67.2
September	56,097	1.7	20.4
October	213,310	6.6	79.2
November	336,914	10.4	124.8
December	343,022	10.5	126.0
Totals	3,252,244	100.0	1200.0

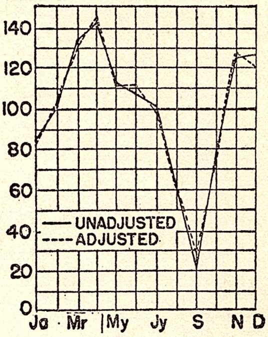


Fig. 27

ADJUSTED INDEX

These index figures do not take into account the differences in the lengths of the months, however. We may adjust our figures, if we are interested in the rate of activity which characterizes the month, by multiplying the actual monthly sales by the adjustment factor shown in Table CI (p. 1256). This table is based on an average of 30.4167 days in each calendar month

$$\left(\frac{365}{12}=30.4167\right)$$

with separate figures for leap years (based on 30.5 days per month). Here it is assumed that every day of the month is of equal importance. There is no allowance for Sundays or holidays. In considering some types of statistics, we should find it wise to work out a similar table, taking the ratio of the number of actual working days in each month to the number of actual working days in the entire year. Again, for a business which, like the amusement industry or the refreshment concession at a pleasure resort, "picks up" on Sundays and holidays, special weighting might be given to the number of special days which occur in a given month. Recognizing, then, the limitations involved in the use of Table CI, we shall employ it in the further illustrations in this section.

We may now adjust the figures in Table 24 by multiplying the actual sales shown opposite each month by the comparable adjustment factors shown in Table CI, getting the results shown in Table 25, column 4. Reducing each of these new figures to percentages of the new total (column 5) and multiplying by 12 (column 6), we obtain an adjusted index which is somewhat more reliable than the index obtained in Table 24. This index, plotted by the dash line in Fig. 27, gives a better picture of the trend for the year under

consideration. Similar computations for all of the years in Table 22 will be necessary before obtaining the over-all trend.

TABLE 25

ADJUSTED INDEX OF FACTORY SALES OF PASSENGER AUTOMOBILES IN THE UNITED STATES—1935, BY MONTHS

Month	Actual Sales	Adjustment Factor	Adjusted Sales Figure	Adjusted Percentage	Adjusted Index
January	227,554	98.11809	223,716	6.9	82.8
February	273,576	108.63071	277,376	8.6	103.2
March	359,410	98.11809	352,646	10.9	130.8
April	387,158	101.38903	392,536	12.1	145.2
May	305,547	98.11809	299,797	9.3	111.6
June	294,182	101.38903	298,260	9.2	110.4
July	274,344	98.11809	269,180	8.3	99.6
August	181,130	98.11809	177,721	5.5	66.0
September	56,097	101.38903	56,876	1.8	21.6
October	213,310	98.11809	209,296	6.5	78.0
November	336,914	101.38903	341,594	10.6	127.2
December	343,022	98.11809	336,567	10.0	120.0
Total	3,252,244		3,235,565	99.7	1196.4

TEST YOUR KNOWLEDGE OF ADJUSTED INDICES

20 Table ε presents the factory sales, by months, of motor trucks and busses in this country for a 10-year period. Find the total sales for each year.

TABLE ε

FACTORY SALES OF MOTOR TRUCKS AND BUSSES IN THE UNITED STATES, 1931-1940

	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1931	33,531	39,521	45,161	50,022	45,688	40,244	34,317	31,772	31,338	21,727	19,683	23,644
1932	20,541	23,308	19,560	27,389	26,539	22,768	14,438	14,418	19,402	13,595	12,025	21 204
1933	18,992	15,319	17,803	26,677	33,760	42,130	38,092	41,441	34,424	29,813	18,318	29,776
1934	42,912	43,482	59,160	64,620	56,691	45,197	41,839	51,311	44,967	47,988	34,462	42,563
1935	62,174	58,655	66,503	65,778	55,560	62,158	57,765	56,270	31,443	58,733	58,145	61,506
1936	66,250	63,331	78,052	86,243	75,591	77,631	68,809	61,923	45,064	34,446	53,902	73,345
1937	70,109	67,405	90,242	96,170	91,487	85,898	78,568	82,874	52,542	31,214	64,727	81,849
1938	53,823	47,151	47,580	43,032	37,101	38,139	34,602	31,870	18,375	22,018	52,069	62,340
1939	60,703	60,220	72,243	63,966	59,672	63,034	58,621	38,461	27,132	61,573	66,533	78,338
1940	69,382	66,276	70,698	70,607	65,539	58,596	62,934	29,050	44,638	72,009	80,261	87,036

21 Compute the percentage which each month's sales is of the annual total, recording the results in tabular form.

22 Table 24 presents the index of monthly sales for the year, 1935. Prepare indices, based on Table 23, for each month of the 10 years.

23 Convert the percentages found in answer to problem 21 into indices, recording the results in a table.

24 Prepare adjusted index figures for each month shown in your answers to problem 22.

25 Adjust the figures found in answer to problem 22.

Secular trend

When we may be contented with a straight-line expression of trend, we may divide our time-series into two equal parts, treating each half separately in determining the averages. (If the whole number of years in such a series is odd, we omit the *middle* year in taking the totals.) We then take on the graph the lines representing the middle year of each half, plot the point for each of the two averages, and connect the two points with a straight line extended the full width of the chart.

SEMI-AVERAGE

The dash line in Fig. 28, superimposed on the curve for the yearly totals, shows us the relationship between yearly sales and the trend of

TABLE 26

COMPUTATION BY SEMI-AVERAGES OF TREND OF FACTORY SALES OF PASSENGER AUTOMOBILES IN THE UNITED STATES, 1931-1940

Year	Sales		Year	Sales
1931	1,973,090		1936	3,669,528
1932	1,135,491		1937	3,915,889
1933	1,573,512		1938	2,000,985
1934	2,177,919		1939	2,866,796
1935	3,252,244		1940	3,692,328
1931-1935	10,112,256	Total	1936-1940	16,145,526
	2,026,128	Average		3,229,105

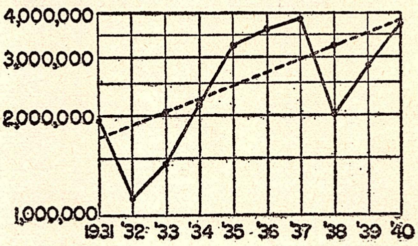


Fig. 28

sales as computed by semi-averages (Table 26). If we have an even number of years in each half, we take as our plotting point the line on the graph which is midway between the two lines representing the two middle years.

This method, it will be noted, does not eliminate all of the cyclical influences, especially when the whole span of the time period is relatively small, but the influence of extremely high or extremely low values is minimized.

MOVING AVERAGE

A better use of arithmetic averages, however, is the computation of "moving" averages. These may be based on the annual totals or on the monthly sales, as need be, and may cover any number of individual items, the choice usually depending on the general impression of trend given by the graph. If progress seems to be upward or downward in three-year cycles, the three-year moving average would be the best guide; if the cycle embraces, on the whole, four or five years, then a four- or a five-year moving average would be computed.

In our illustration, Table 27, we begin by computing the "moving total" for each two years in the list. Since sales in 1931 amounted to 1,973,090 and

TABLE 27

COMPUTATION OF MOVING ANNUAL AVERAGES OF FACTORY SALES OF PASSENGER AUTOMOBILES IN THE UNITED STATES, 1931-1940

		—Two Year—		—Three Year—		—Four Year	
Year	Sales	Moving Total	Moving Average	Moving Total	Moving Average	Moving Total	Moving Average
1931	1,973,090	3,108,581	1,554,290.5				
1932	1,135,491	2,709,003	1,354,501.5	4,682,093	1,560,697.67	6,860,012	1,715,003.00
1933	1,573,512	3,751,431	1,875,715.5	4,886,922	1,628,974.00	8,139,166	2,034,791.50
1934	2,177,919	5,430,163	2,715,081.5	7,003,675	2,334,558.33	10,673,203	2,668,300.75
1935	3,252,244	6,921,772	3,460,886.0	9,099,691	3,033,230.33	13,015,580	3,253,895.00
1936	3,669,528	7,585,417	3,792,708.5	10,837,661	3,612,553.67	12,838,646	3,209,661.50
1937	3,915,889	5,916,874	2,958,437.0	9,586,402	3,195,467.33	12,453,378	3,113,344.50
1938	2,000,985	4,867,781	2,433,890.5	8,783,670	2,927,890.00	12,475,998	3,118,999.50
1939	2,866,796	6,559,124	3,279,562	8,560,109	2,853,369.00		
1940	3,692,328						

sales in 1932 amounted to 1,135,491, our total for the two years would be 3,108,581. This figure is entered in column 3 of Table 27 on a line midway

between the 1931 figures and the 1932 figures. Dividing this figure by 2 enables us to determine the average for the 2 years, as recorded in column 4. We proceed in the same way with the totals for 1932 and 1933, and so on down the column.

The three-year columns are computed in a similar manner, the figures opposite 1932 in column 5 being the arithmetic mean of the totals for 1931, 1932, 1933, etc. In the same way, the figures between 1932 and 1933 in the four-year columns represent the total and averages of 1931, 1932, 1933, 1934, and so on.

Fig. 29 shows the moving averages for two-, three-, and four-year periods plotted on the same coördinates graph with the figures for the single years. Here it is readily noted that the curve tends to smooth out in proportion to the span of the average. Too great a span would nullify the usefulness of the computation. Here the factor of judgment must always enter in to determine what span is proper for indicating the exact facts under consideration.

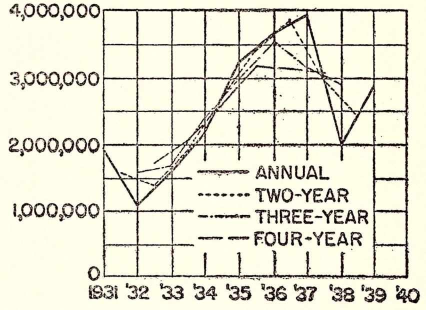


Fig. 29

The same sort of moving average may be applied to the monthly sales figures. In this case, the span may even be extended to 12, so that we get 12 annual totals for each year (January to December, February to the next January, March to the next February, etc.)

Table 28 shows this application to the figures for the period, July, 1938, through June, 1940. The 12-months' total shown in column 3 above January is for the period from July, 1938, to June, 1939; that between January and February is for the period from August, 1938, to July, 1939; etc. Column 4 presents the averages for the same 12-months' periods. From these two averages, we obtain the centered average shown opposite January. The

TABLE 28

COMPUTATION OF MOVING AVERAGES OF FACTORY SALES OF PASSENGER AUTOMOBILES IN THE UNITED STATES—1939

Month	Sales 1939	12-Months' Total	12-Months' Average	2-Months' Moving Total	Centered 12-Months' Average	Ratio
		2,646,619	220,552			
January	281,465			444,762	222,381	126.6
		2,690,516	224,210			
February	243,000			448,652	224,326	108.3
		2,693,299	224,442			
March	299,703			456,922	228,461	131.1
		2,789,765	232,480			
April	273,409			470,321	235,161	112.0
		2,854,090	237,841			
May	237,870			472,758	236,379	100.6
		2,818,998	234,917			
June	246,704			473,817	236,909	104.1
		2,866,796	238,900			
July	150,738			484,586	242,293	62.2
		2,948,228	245,686			
August	61,407			499,268	249,634	24.9
		3,042,984	253,582			
September	161,625			511,599	255,799	63.2
		3,096,203	258,017			
October	251,819			523,428	261,714	96.2
		3,184,933	265,411			
November	285,252			538,139	269,069	106.1
		3,272,739	272,728			
December	373,804			548,734	274,367	136.3
		3,312,075	276,006			

January ratio is determined by dividing the actual January sales (column 2) by the centered average for January (column 6) and recorded in column 7. We proceed in the same way to December, where column 3 shows the 12-months' total for the period from June, 1939, to May, 1940. The 12-months' averages shown in column 4 are easily computed by dividing each of the totals in column 3 by 12.

Since we are dealing here with the arithmetic mean of an even number of items, it is evident that our 12-months' averages describe in each case a spot between two months. We overcome this difficulty by *centering* the averages. To accomplish this, we take each pair of consecutive averages, add them, and divide by 2. These results are recorded in column 6. The ratios in column 7 are obtained by dividing the figures in column 2 by the corresponding figures in column 6.

TEST YOUR KNOWLEDGE OF SEMI-AVERAGES AND MOVING AVERAGES

26 Compute the trend shown by the figures in Table ε by the method of semi-averages.
27 Compute the trend for the same figures by the method of moving annual averages (a) for a two-year, (b) for a three-year, (c) for a four-year period; and record your data as in Fig. 29.
28 Using Table 22, compute the ratios for each month of the 10 years, using the method of Table 28 and recording the data as in Table 23.
29 Using Table ε, compute the monthly ratios, as in problem 28.

THE LEAST-SQUARES METHOD

To secure the line of best fit in estimating trend, we employ the least-squares method, computed in such a way that the sum of the squared deviations of the observed values about the trend line is a minimum. The formula for the straight line is written:

$$y_c = a + bx, \qquad \text{XXVI}$$

where

y_c = the computed values of the trend;
a = value of the line at its origin;
b = the typical increase or decrease occurring in x units of time;
and x = the time interval (*e.g.*, expressed as 1 month, 6 months, or 1 year).

Summing up the points in XXVI gives us

$$\Sigma y = \Sigma a + b\Sigma x,$$

or, since $\Sigma a = na$,

$$\Sigma y = na + b\Sigma x. \qquad \text{XXVIIa}$$

Multiplying every term in XXVI by x (the coefficient of b) gives us

$$xy = ax + bx^2,$$

which, summed, becomes

$$\Sigma(xy) = a\Sigma x + b\Sigma(x^2). \qquad \text{XXVIIIa}$$

We now have, in XXVIIa and XXVIIIa, two simultaneous equations which may readily be solved (cf. pp. 137-141).

Returning to the yearly totals from Table 22, we reproduce them in column 2 of Table 29. Letting the first year, 1931, equal 0, we should designate the next year, 1932, as 1, and so on, as shown in column 3. This is our x value. The values of x^2, then, will be as indicated in column 4. The xy values in column 5 are obtained by finding the products of the pairs of values in columns 2 and 3.

TABLE 29

COMPUTATION OF TREND OF FACTORY SALES OF PASSENGER AUTOMOBILES IN THE UNITED STATES, 1931-1940—LEAST-SQUARES METHOD

Year	Actual Sales y	Deviation x	Square of Deviation x^2	Product xy	Ordinates of Trend
1931	1,973,090	0	0	0	1,661,430
1932	1,135,491	1	1	1,135,491	1,875,730
1933	1,573,512	2	4	3,147,024	2,090,029
1934	2,177,919	3	9	6,533,757	2,304,329
1935	3,252,244	4	16	13,008,976	2,518,628
1936	3,669,528	5	25	18,347,640	2,732,928
1937	3,915,889	6	36	23,495,334	2,947,228
1938	2,000,985	7	49	14,006,195	3,161,527
1939	2,866,796	8	64	22,934,368	3,375,827
1940	3,692,328	9	81	33,230,952	3,590,126
$n=10$	$\Sigma y=26{,}257{,}782$	$\Sigma x=45$	$\Sigma(x^2)=285$	$\Sigma(xy)=135{,}839{,}737$	

To obtain the values in column 6, we substitute in the formulas values previously found in earlier columns of the table.

Then, since

$$\Sigma y=26{,}257{,}782, \qquad n=10 \qquad \text{and} \qquad \Sigma x=45,$$

formula XXVIIa becomes

$$\Sigma y=na+b\Sigma x$$
$$26{,}257{,}782=10a+45b \qquad \text{a}$$

and formula XXVIIb becomes

$$\Sigma xy=a\Sigma x+b\Sigma x^2$$
$$135{,}839{,}737=45a+285b. \qquad \text{b}$$

Multiplying equation a by 9 and equation b by 2, and subtracting, we have

$$\begin{aligned} 236{,}320{,}038&=90a+405b\\ 271{,}679{,}474&=90a+570b\\ \hline -35{,}359{,}436&=\quad -165b\\ b&=214{,}299.6 \end{aligned}$$

Substituting this value in equation a, we have

$$\begin{aligned} 26{,}257{,}782&=10a+45\cdot 238{,}542\\ 26{,}257{,}782&=10a+10{,}734{,}390\\ 10a&=15{,}523{,}392\\ a&=1{,}552{,}339. \end{aligned}$$

Then equation XXVI becomes

$$y_c=1{,}661{,}430+214{,}299.6x. \qquad \text{c}$$

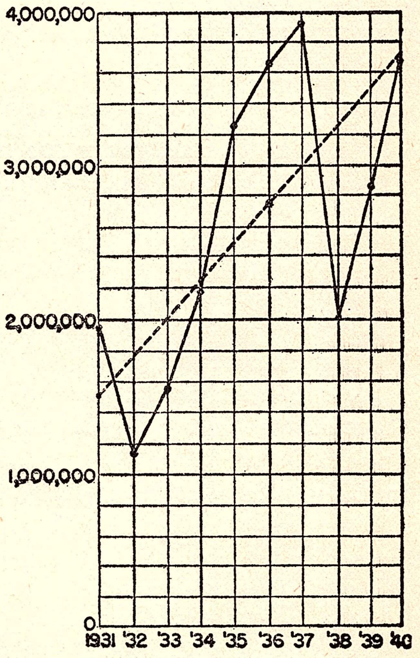


Fig. 30

To obtain the various trend values for y, we substitute the various values of x (shown in Table 29) in this equation, getting the results which are recorded in column 6 of Table 29. Since this is a straight-line trend, we really need to compute only the values for any two years in order to obtain the slope of the line.

Since $x=0$ for 1931, we get the ordinate of trend for that year by taking merely the first figure, since

$$y_{c1931}=1{,}661{,}430+214{,}299.6\cdot 0=1{,}661{,}430+0=1{,}661{,}430.$$

For 1936, our computation, by a similar line of reasoning, would be

$$y_{c1936}=1{,}661{,}430+214{,}299.6\cdot 5=1{,}661{,}430+1{,}071{,}498=2{,}732{,}928.$$

The other values we may read approximately from the graph in Fig. 30 or we may compute each in turn by further substitution in equation c.

Comparing column 6 with column 2, we note that the sales in 1932 were actually considerably less than the trend ordinate for that year calls for. Since we know that 1932 was a "depression year", we are not surprised at this finding. We see, too, that the figures for the years, 1933 to 1937, inclusive, are considerably above the trend line, that the figures for 1938 and 1939 again drop below it, but that the 1940 figure brings the sales almost back to normal. This is in line with our general knowledge of the "second depression" or "regression" which occurred in the United States at about that time.

Short method, when number of items is odd—When we are dealing with an odd number of items, we may shorten the process considerably by taking the median year as the origin, counting both ways from this point to determine the values of x, as in Table 30.

TABLE 30

COMPUTATION OF TREND OF FACTORY SALES OF PASSENGER AUTOMOBILES IN THE UNITED STATES, 1931-1939—LEAST-SQUARES METHOD—SHORT-CUT FOR ODD NUMBER OF YEARS

YEAR	ACTUAL SALES	DEVIATION	SQUARE OF DEVIATION	PRODUCT	ORDINATES OF TREND
	y	x	x^2	xy	
1931	1,973,090	−4	16	−7,892,360	1,684,109
1932	1,135,491	−3	9	−3,406,473	1,889,900
1933	1,573,512	−2	4	−3,147,024	2,095,691
1934	2,177,919	−1	1	−2,177,919	2,301,482
1935	3,252,244	0	0	0	2,507,273
1936	3,669,528	1	1	3,669,528	2,713,064
1937	3,915,889	2	4	7,831,778	2,918,855
1938	2,000,985	3	9	6,002,755	3,124,646
1939	2,866,796	4	16	11,467,184	3,330,437
$n=9$	$\Sigma y=22{,}565{,}454$	$\Sigma x=0$	$\Sigma(x^2)=60$	$\Sigma(xy)=12{,}347{,}469$	

Since $\Sigma x=0$, $b\Sigma x$ will also equal 0 and formula XXVIIa becomes

$$\Sigma y=na. \qquad \textbf{XXVIIb}$$

Similarly, $a\Sigma x$ will be equal to 0 and formula XXVIIIa becomes

$$\Sigma(xy)=b\Sigma(x^2). \qquad \textbf{XXVIIIb}$$

We no longer need to solve the equations simultaneously. Our unknown a is found directly by substitution in XXVIIb while b is found in like fashion from XXVIIIb.

From the totals in Table 30, we note

$$\Sigma y=22{,}565{,}454, \quad n=9, \quad \Sigma(xy)=12{,}347{,}469, \text{ and } \Sigma(x^2)=60.$$

Then, by XXVIIb, $22{,}565{,}454=9a$

$$a=2{,}507{,}273,$$

and, by XXVIIIb, $12{,}347{,}469=60b$

$$b=205{,}791.$$

Our equation, based on formula XXVI, then becomes

$$y_c=2{,}507{,}273+205{,}791x,$$

from which we easily compute the values for column 6 of Table 30.

Short method, when number of items is even—If the whole number of items is even, we can still use the short method, with a slight difference. We take the origin at a point midway between the two middle years, computing our x values for these years as -0.5 and $+0.5$, respectively, and again moving to the ends of the scale by subtracting or adding 1 to indicate the distance moved each year, as in column 3 of Table 31. As these x values would entail additional labor no matter

TABLE 31

COMPUTATION OF TREND OF FACTORY SALES OF PASSENGER AUTOMOBILES IN THE UNITED STATES, 1931-1940—LEAST-SQUARES METHOD—SHORT-CUT FOR EVEN NUMBER OF YEARS

Year	Actual Sales	Deviations in Years	Deviations in Half-Years	Deviations in Half-Years Squared	Product	Ordinates of Trend
	y	x	x'	$(x')^2$	$x'y$	
1931	1,973,090	−4.5	−9	81	−17,757,810	1,661,392
1932	1,135,491	−3.5	−7	49	−7,978,437	1,875,700
1933	1,573,512	−2.5	−5	25	− 7,867,560	2,090,008
1934	2,177,919	−1.5	−3	9	−6,533,757	2,304,316
1935	3,252,244	−0.5	−1	1	−3,252,244	2,518,624
1936	3,669,528	0.5	1	1	3,669,528	2,732,932
1937	3,915,889	1.5	3	9	11,747,667	2,947,240
1938	2,000,985	2.5	5	25	10,004,925	3,161,548
1939	2,866,796	3.5	7	49	20,067,572	3,375,856
1940	3,692,328	4.5	9	81	33,230,952	3,590,164
$n=10$	$\Sigma y=26,257,782$	$\Sigma x=0$	$\Sigma(x')=0$	$\Sigma(x')^2=330$	$\Sigma(x'y)=35,360,836$	

whether we used them as decimal or as common fractions, we simplify our problem by considering the number of half years instead of the number of years, using x' (x prime) to designate this new value, as in column 4. Squaring the x' values, we get the results shown in column 5. Multiplying the y values in column 2 by the x' values in column 4, we get the results recorded in column 6.

From the totals line, we see that

$$\Sigma y=26,257,782, \quad n=10, \quad \Sigma(x'y)=35,360,836, \quad \text{and } \Sigma(x')^2=330.$$

Then, by XXVIIb, $\quad 26,257,782=10a$

$$a=2,625,778$$

and, by XXVIIIb, $\quad 35,360,836=330b$

$$b=107,154.$$

Here, however, b is the coefficient of x', which measures half-year values, and therefore represents the increase for each half year. In computing the estimates for each year, we multiply 109,972 by the value in the x' column or we double the amount and multiply it by the value in the x column.

$$y_{c1936}=2,625,778+107,154$$

or $\quad y_{c1936}=2,625,778+214,308\cdot 0.5,$

the answer in either case being 2,732,932.

Similarly,

$$y_{c1940}=2,625,778+214,308\cdot 4.5$$

or $\quad y_{c1940}=2,625,778+107,154\cdot 9=2,625,778+964,386=3,590,164,$

and so on for each of the other years, as recorded in column 7 of Table 31.

Comparison of these figures with the figures in Table 29 shows slight but not significant differences. The time saved in computation more than offsets the slight variance in degree of accuracy, particularly since in both cases the figures are merely estimates.

TEST YOUR KNOWLEDGE OF THE LEAST-SQUARES METHOD

30 Using the least-squares method, compute the annual trend for the data in Table ε.

31 Use the short-cut method (least squares) for computing the annual trend of the same data. Compare your results with the answers to problem 30.

32 (a) Eliminate the figures for 1931 and then compute (by the short method) the annual trend for the remaining data in Table ε. (b) Restore the figures for 1931 and eliminate the figures for 1940. Again use the short-cut method to compute the annual trend for the remaining data.

GEOMETRIC STRAIGHT-LINE TREND

Formulas XXVIIa and XXVIIIa may be utilized for computing the geometric straight-line trend merely by using them in the logarithmic form:

$$\Sigma \log y = n \log a + \log b \, \Sigma x \qquad \textbf{XXVIIc}$$

$$\Sigma(x \log y) = \log a \, \Sigma x + \log b \, \Sigma(x^2) \qquad \textbf{XXVIIIc}$$

Similarly, when the short-cut method is employed, formulas XXVIIb and XXVIIIb may be expressed logarithmically as:

$$\log a = \frac{\Sigma \log y}{n} \qquad \textbf{XXVIId}$$

$$\log b = \frac{\Sigma(x \log y)}{\Sigma(x^2)} \, . \qquad \textbf{XXVIIId}$$

From these,

$$\log y_c = \log \mathrm{a} + (\log b)x \qquad \textbf{XXIXa}$$

or

$$y_c = a(b)^x \, . \qquad \textbf{XXIXb}$$

CURVED-LINE TRENDS

Other trend lines which are frequently employed in statistical work present additional advantages. To round out the discussion, we present them although space does not permit illustrative examples.

The second-degree parabola, expressed by

$$y_c = a + bx + c^2 \, , \qquad \textbf{XXX}$$

is determined by solving these equations:

$$\Sigma y = na + b\Sigma x + c\Sigma x^2$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$$
$$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4,$$

or, in the case of the short method,

$$\Sigma y = na + c\Sigma x^2$$
$$\Sigma xy = b\Sigma x^2$$
$$\Sigma x^2 y = a\Sigma x^2 + c\Sigma x^4.$$

Here the b and the c values counteract each other, so that the curve has both a positive and a negative stage.

The third-degree parabola is similar:

$$y_c = a + bx + cx^2 + dx^3. \qquad \text{XXXI}$$

It is obtained by solving a group of four equations:

$$\begin{aligned} \Sigma y &= na + b\Sigma x + c\Sigma x^2 + d\Sigma x^3 \\ \Sigma xy &= a\Sigma x + b\Sigma x^2 + c\Sigma x^3 + d\Sigma x^4 \\ \Sigma x^2y &= a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 + d\Sigma x^5 \\ \Sigma x^3y &= a\Sigma x^3 + b\Sigma x^4 + c\Sigma x^5 + d\Sigma x^6. \end{aligned}$$

Again, when the short method is used, these equations become simpler:

$$\begin{aligned} \Sigma y &= na + c\Sigma x^2 \\ \Sigma xy &= b\Sigma x^2 + d\Sigma x^4 \\ \Sigma x^2y &= a\Sigma x^2 + c\Sigma x^4 \\ \Sigma x^3y &= b\Sigma x^4 + d\Sigma x^6. \end{aligned}$$

Here the trend changes direction twice.

The Gompertz curve, recommended for use in analyzing the secular movement in industry, is obtained from the formula,

$$y_c = ab^{c^x} \qquad \text{XXXIIa}$$

expressed logarithmically as

$$\log y_c = \log a + (\log b)c^x. \qquad \text{XXXIIb}$$

The logistic curve, also known as the Pearl-Reed curve, is useful when the data do not increase at a rate faster than the geometric rate; if the data increase faster, they fall outside the range of this curve. Its formula is

$$\frac{1}{y_c} = a + bc^x. \qquad \text{XXXIII}$$

TEST YOUR KNOWLEDGE OF OTHER TREND LINES

33 Prepare a table showing computation of the geometric straight-line trend for data of Table 22 from 1932 to 1940. Column headings: 1, year; 2, sales (the y value); 3, $\log y$; 4, x (counting the middle year as 0, as in Table 30); 5, x^2; 6, $x \log y$; 7, $\log a + (\log b)x$; 8, anti-logarithms of figures in column 7.

34 Record on a graph the results found in problem 33. Compare your readings from the trend line with your computations.

35 Compute the geometric straight-line trend for data of Table ε, 1931 to 1939.

36 Use formulas XXX and XXXI for computing second-degree parabolas for data of Tables 22 and ε.

Cyclical variation

We may reduce to formula the relationships among the movements mentioned on page 1216 by letting: T = trend, S = seasonal variation, C = cyclical movement, I = irregular variations, and O = original data.

Then $$O = T \times S \times C \times I. \qquad \text{XXXIVa}$$

When we analyze annual totals, the movement may be expressed as

$$O=T\times C\times I, \qquad \text{XXXIVb}$$

since the effect of seasonal variations disappears.

To separate the secular movement from the cyclical and irregular variations, we merely divide XXXIVb by T:

$$\frac{O}{T}=\frac{T\times C\times I}{T}=C\times I. \qquad \text{XXXIVc}$$

Repeating columns 2 and 7 from Table 31 as columns 2 and 3 of Table 32, we obtain the figures for column 4 of Table 32 by dividing each item from column 2 by the corresponding item in column 3.

TABLE 32

ADJUSTED INDICES OF ANNUAL FACTORY SALES OF PASSENGER AUTOMOBILES IN THE UNITED STATES—1931-1940

YEAR	ACTUAL SALES	ORDINATES OF TREND	SALES ADJUSTED FOR TREND
1931	1,973,090	1,661,392	118.7
1932	1,135,491	1,875,700	60.5
1933	1,573,512	2,090,008	75.3
1934	2,177,919	2,304,316	94.5
1935	3,252,244	2,518,624	129.1
1936	3,669,528	2,732,932	134.3
1937	3,915,889	2,947,240	132.8
1938	2,000,985	3,161,548	63.3
1939	2,866,796	3,375,856	84.9
1940	3,692,328	3,590,164	102.8

Fig. 31, in which the trend index (column 4 of Table 32) is plotted, should now be compared with Fig. 30, in which the trend line is shown. Since we have now reduced the original items to percentages of trend, the line of comparison becomes straight and horizontal, while the deviations are noted by the curve which rises and falls in relation to it. This straight line is known as the *statistical normal* and the values as plotted on the Y-axis are taken as describing the per cent of normal attained each year.

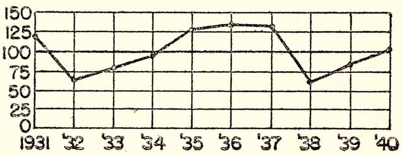


Fig. 31

If it is desired to analyze monthly data in this fashion, it is necessary to determine the trend ordinate for each month, multiply this by the seasonal index, and use the product as a divisor for the original figure for that month. Plotting of this adjusted production index gives a picture of cyclical and random variations. This curve may be smoothed by taking a moving average (cf. p. 1222), thus providing an estimate of the cyclical influence.

In comparing two curves, we first obtain the deviation of each point from the normal—that is, we subtract 100 from each index. Then we

calculate the *standard deviation of deviations* from the normal, using the formula,

$$\sigma = \sqrt{\frac{\Sigma(d^2)}{n} - \left(\frac{\Sigma d}{n}\right)^2}. \qquad \text{XXXV}$$

Dividing the deviation from the normal by the standard deviation, we have an estimate of the outside limits within which our computation will be valid.

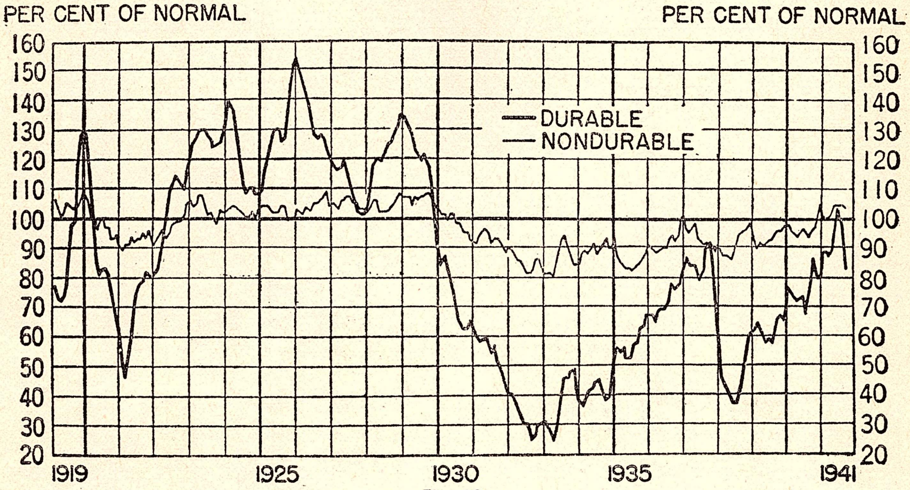


Fig. 32

Fig. 32 shows the per cent of normal for each year's production of consumers' durable and nondurable goods in the United States from

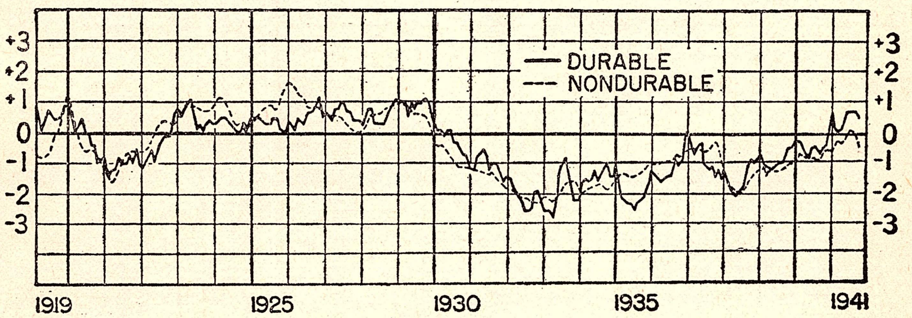


Fig. 33

1919 to 1941 inclusive, while Fig. 33 shows the standard deviation scale for the same figures.

TEST YOUR KNOWLEDGE OF TREND ADJUSTMENT AND STANDARD DEVIATION

37 Adjust for trend the annual sales of motor trucks and busses, using data of Table ε.

38 Compute the standard deviation of deviations of data from Table 22 and ε.

SAMPLING When we are dealing with statistical analysis, it is not always possible for us to get all of the data which we might like to use. Even the United States census, thorough-going as it is, probably misses some cases through the inability of census-takers to contact persons who are in process of moving from place to place during the time the census is under way. How much harder it would be for an individual to track down every last case when dealing with a huge mass of data!

In other instances, the mass of data at hand is so tremendously large that the computations involved in handling it would be too severe a tax on time and energy. For that reason, we resort to *sampling*, by which we mean the selection of representative data upon which we may base conclusions as to the probable nature of the whole.

Following the line of reasoning concerning the normal curve (discussed on page 1188), we secure the curve for that portion of the data which we do submit to analysis and upon our results reach conclusions as to the likelihood that the findings will apply to the whole. This is the theory on which "straw votes" and such fact-finding investigations as the Gallup Poll or the National Opinions Poll are based.

Selecting the sample

Obviously, one who is bent on "proving" a certain point might, by selecting the samples with an eye to the results, so distort his figures as to appear to have ascertained certain "facts which just aren't true". Just as obviously, a really conscientious statistician who happened to overlook a chance factor might deceive himself as well as his public (as a well-known magazine did a few years ago) as to the reliability of his findings.

CRITERIA FOR SAMPLING

To avoid such possibilities, we should, when using the sampling process, set up criteria by which to determine the validity of the operations. To this end, we must be certain that:

a Our sample has been selected in such a way that there is no prejudice or bias in the figures. (By interviewing only red-headed candidates for a secretarial position, we might arrive at the statement that all stenographers have red hair.)

b There is complete independence among the items or cases selected. (By interviewing only wives of men with small incomes, we might arrive at the conclusion that no woman is in a position to buy an expensive mink coat or to spend more than a certain amount on the family's food bill.)

c The sample represents the entire area to which the conclusions are to be applied. (In trying to determine living conditions throughout the United States, we should get wholly different pictures if we limited our samples

to the Kentucky mountains, to a New England agricultural area, or to the suburban villages of New York City.)

d Conditions are the same for all items included in the sample.

TYPES OF SAMPLES

We may arrive at our sample by any one of several different methods, the choice depending upon the exact purpose we have in mind.

The random sample is used when we desire to get a general picture of conditions over the whole area under investigation. Having collected our data, we might write each item on a separate card, shuffle the whole pack, and then draw out a number of cards sufficient to serve as our sample. For even greater accuracy, we should re-insert each card and re-shuffle the stack after each drawing. To most people, the word, *random*, implies an entirely accidental selection. For statistical purposes, such a random sample would be of little use to us, since we might entirely miss the extremes and so reach an utterly false conclusion.

A controlled random sample in which the selection is made in such a way as to give each item the same chance of being included, overcomes this difficulty and adds to the statistical worth of our findings. To assure proper representation of all sections of the total population (the whole number of cases to be found for the subject under consideration), the investigator first determines the approximate percentage which each group is of the whole group and then is careful to see that his sample includes a proportionate number from each of the groups.

Such a sample is frequently referred to as a *stratified random sample* because the population is divided into *strata*, or layers. We might desire to determine the shades of opinion or amount of income, for example, of persons in various geographical divisions of the nation, or we might desire to ascertain separately views or preferences of persons of differing political allegiances, or to arrange our cases according to income or age groups. Unless we have first made certain that our sample is a true sample in that its findings would conceivably hold true if we included every possible case, we are in danger of distorting our conclusions because of the bias introduced into the sample.

Purposive samples, such as the controlled and stratified samples, are statistically reliable, within the limits which will be considered later, if the purpose behind their selection is in line with the criteria we have already set up.

DETERMINING THE ADEQUACY OF THE SAMPLE

In line with our previous knowledge of the curve of normal distribution (p. 1188), we know that, the greater the number of cases, the nearer we shall expect to approach the normal curve in our findings. As the size of the sample becomes larger, we find a decrease in the size of the *sampling error* (the amount by which the figure arrived at by computing from the data of the sample is likely to differ from the actual facts). The mathematical principle by which to determine the amount by which the error is reduced is stated simply: the extent of the sampling error decreases as the square root of the number of items in the sample

increases; that is, in order to cut the sampling error in half, we must increase the size of the sample four times; to reduce the error to one-third of its former size, we must take a sample nine times as large as the original sample; etc.

A rule-of-thumb method of determining the adequacy of a sample is to tabulate and make computations for one portion of the sample, then of another, comparing the results of the second with those of the first, and then comparing the two sets of results with the total results; continuing to add portions of somewhat similar size until the amount of variance in the results is so slight as to indicate that addition of still further portions would not be likely to change the picture to any great extent.*

Adequacy, then, is something more than size. In some instances, 30 to 50 well-selected cases will give a more nearly correct picture than will thousands of cases ill-advisedly selected.

Determining extent of error

When discussing *error* in connection with sampling, we are using the word in a specialized sense to indicate the amount by which the findings based on the sample are likely to deviate from the findings which we should reach if it were possible to investigate and include in our computations all of the existing cases. Having made calculations based on a given sample, we are interested in determining how accurate the findings are likely to be in giving a picture of the total situation. To do this, we may calculate either the standard error or the probable error.

STANDARD ERROR OF THE MEAN

Since the standard error implies a deviation, we use as our symbol for standard error the same symbol (σ) which we have already used for standard deviation (p. 1210). When discussing standard error, however, we annex to this symbol a subscript indicating the measure for which we are determining the error. In the case of the standard error of the arithmetic mean, then, we have σ_{M_A}. Similarly, using S to indicate "sample", we achieve as our formula for standard error of the mean:

$$\sigma_{M_A} = \frac{\sigma_S}{\sqrt{n-1}}. \qquad \text{XXXVI}$$

Assuming that a sufficiently large number of samples has been taken, we should find that the arithmetic mean would divide the normal curve into exact halves. When we mark off a range of one standard error on either side of this point, as in Fig. 34, we have included 68.27 per

* *A word of caution*: no matter how large the sample, errors due to bias will not be reduced. For this reason, care must be taken in the first place to make sure that the cases selected for the sample are truly representative.

cent of the sample; marking off two standard errors each way brings us to 95.45 per cent of the cases, while all but 0.27 per cent will be included if we increase our range to three standard errors in each direction. The chances, then, are only about 45 in one thousand that a difference as great as 2 σ_{M_A} could have occurred as a result of random errors. (For further discussion of probability see Table CII, page 1257.)

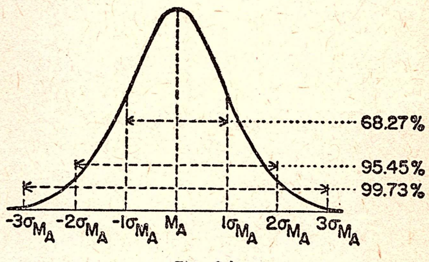


Fig. 34

Taking the 144 sales recorded in Table 1, we recognize that the next group of 144 sales or, for that matter, any other group of 144 sales, would probably not reach exactly the same total. However, these other groups of sales might be such that this first group would be a fairly representative sample. That is what we are interested in ascertaining. We have already discovered (p. 1211) that ± 0.51 is the standard deviation for the mean, \$2.50. Substituting this figure in formula XXXVIa, we have

$$\sigma_{M_A} = \frac{0.51}{\sqrt{144-1}} = \frac{0.51}{11.958} = 0.0426.$$

In giving a report concerning the arithmetic mean of the first 144 sales as a sampling investigation of sales conditions, then, we should state that the mean of such sales is \$2.50±\$0.0426. In every report of a sampling investigation, the measure should be stated in this fashion so that the reader may have an opportunity to judge for himself as to the reliability of the computed measure in representing the probable true measure.

Small samples—When the values of a measure are computed from a small sample (30 cases or less) or even from a large number of small samples, the distribution of these values is not necessarily normal. To test the significance of deductions based on a small sample, we may use

$$t = \frac{k_S - k_U}{\frac{\sigma}{n-m+1}} \qquad \text{XXXVII}$$

where k_S = value of any given measure in the sample,
k_U = value of the same measure in the universe,
n = number of items in the sample,
and m = number of predetermined conditions in the sample.

We may also make use of Table CIII (p. 1257), in which we are provided the multiples by which the mean is to be amplified in order to attain the indicated percentages of probability of the occurrence of a deviation.

The size of this critical ratio is interpreted by reference to Fig. 34, just as in the case of standard error.

Differences between sample means—In the case of using several different samples, to determine the significance of the difference between two means, we use σ_D to indicate the standard error of the differences between paired means.

$$\sigma_{D_{1-2}}\sqrt{\sigma^2_{M_{A_1}}+\sigma^2_{M_{A_2}}}. \qquad \text{XXXVIII}$$

The critical ratio of this is

$$T=\frac{M_{A_1}-M_{A_2}}{\sigma_{D_{1-2}}}. \qquad \text{XXXIX}$$

STANDARD ERROR OF OTHER MEASURES OF CENTRAL TENDENCY

In similar fashion, we may determine the standard error of any other of the measures of central tendency. The formulas for these are given in column 3, Fig. 35. Since the process of using these formulas is similar to the process for formula XXXVI, we leave the computations to the reader.

PROBABLE ERROR

Instead of standard error, some statisticians make use of the probable error. This is the figure which will not be exceeded by 50 per cent of the cases. It may be computed by multiplying the standard error by 0.6745, as shown in column 4, Fig. 35. In general, the standard error is to be preferred to the probable error in statistical work.

INTERPRETATION OF SAMPLING

Measure	Formula for Measure	Formula for Standard Error of Measure	Formula for Probable Error of Measure*
Arithmetic mean	$M_A=\frac{\Sigma v}{n}$	$\sigma_{M_A}=\frac{\sigma_S}{\sqrt{n-1}}$ †	$\text{PE}_{M_A}=0.6745\cdot\frac{\sigma}{\sqrt{n-1}}$
Median	$M_d=\frac{n+1}{2}$	$\sigma_{M_D}=1.2530\cdot\frac{\sigma}{\sqrt{n}}$ ‡	$\text{PE}_{M_d}=0.84535\cdot\frac{\sigma}{\sqrt{n}}$
Standard deviation	$\sigma=\sqrt{\frac{\Sigma(d_n)^2}{n}}$	$\sigma_\sigma=\frac{\sigma}{\sqrt{2n}}$ ‡	$\text{PE}_\sigma=0.6745\cdot\frac{\sigma}{\sqrt{2n}}$
	$\sigma=i\sqrt{\frac{\Sigma[f(d_m)^2]}{\Sigma f}-\left(\frac{\Sigma fd_m}{\Sigma f}\right)^2}$		
Coefficient of variation	$V=\frac{\sigma}{M_A}\cdot 100$	$\sigma_V=\frac{V}{\sqrt{2n}}\sqrt{1+\frac{2(V^2)}{10^4}}$	$\text{PE}_V=0.6745\cdot\frac{V}{\sqrt{2n}}\sqrt{1+\frac{2(V^2)}{10_4}}$

Fig. 35

* See Tables CIV to CVI, pp. 1258 to 1260.
† Universe standard deviation not known.
‡ Universe standard deviation known.

INDEX NUMBERS In recent years, it has become increasingly common to try to indicate changes in data by expressing current figures as an index. This index is computed in much the same way as was the index of sales figures (cf. p. 1220), but here the divisor is determined in a different fashion. For the base period, we select a year (or the average of several years) in the not-too-distant past which may be considered normal. Then we compare the current figures with this norm, or base, by dividing each of them by it.

It should be obvious that wholly different results will be obtained if a different year is selected as the base, so that a picture of prosperity or the reverse may be "proved" by shifting the base from one point to another. In general, the index should be computed on a base which is itself the average of a sufficient number of items to ensure the elimination, to a reasonable degree, of chance factors. Nothing less than 20 items should be considered satisfactory, and 50 is a much safer guide. Past 50, the increases in reliability are slight; past 200, according to some authorities, there is no perceptible gain.

Illustrative Example

Table 33 shows some actual figures for a fairly long interval of years. In successive columns, as shown by the headings, we should get wholly different index values if we used the indicated bases.

TABLE 33

INDICES COMPUTED ON VARIOUS BASES

Year	Amount	1925 Base	1932 Base	1914 Base (6050)	1914-19 Base (Av. =9835)	1935-39 Base (Av. =8472)
1925	10,995	100.0	231.8	186.6	111.1	129.7
1926	10.564	96.1	222.7	174.6	107.4	124.7
1927	10,756	97.8	226.7	177.7	109.3	126.9
1928	11,072	100.7	233.4	183.0	112.5	130.8
1929	11,296	102.7	230.5	186.7	114.6	133.3
1930	9,021	82.5	190.2	149.1	91.6	106.5
1931	6,371	57.9	134.3	105.3	64.8	75.2
1932	4,743	43.1	100.0	78.4	48.2	55.9
1933	5,445	49.5	114.8	90.0	55.4	64.3
1934	6,780	61.7	142.9	112.1	68.9	80.0
1935	7,659	69.7	161.4	126.6	77.8	90.4
1936	8,654	78.7	184.6	143.0	87.9	102.1
1937	9.217	83.8	194.3	152.3	93.6	108.8
1938	8,168	74.3	172.2	135.0	83.5	96.4
1939	8.684	78.9	183.1	143.5	88.3	102.5
1940	9,145	83.2	192.7	151.1	92.9	107.9
1941	11,830	107.6	247.3	195.5	120.3	139.6

From this, using the 1925 base as 100, we find that conditions in 1941 were 7.6 per cent better, whereas, if we use a 1914 base (which was at one time the customary procedure), we should find conditions 95.5 per cent better. Using the depression year of 1932 as a base would show a 1941 improvement of 147.3 per cent. If we took the average of the years, 1914 to 1919 (these figures are not shown in the table), we should find that it was 9835 and that conditions in 1941 showed an improvement of 20.3 per cent. Coming closer to the present and avoiding the depression of the early 1930's, we might take 1934 to

1939 as the base, getting the average, 8472, and showing an improvement in 1941 of 39.6 per cent.

These calculations show the necessity in all reports on index numbers of indicating how the base has been determined. They will also serve to explain how rival groups may arrive at different findings and cite apparently convincing figures to "prove" their contentions. This typical passage is quoted from a report on indices of wholesale, retail, and farm prices*:

The wholesale-price index of the Department of Labor is based on primary market quotations of 784 commodities beginning 1926, 813 beginning January 1938, 863 beginning March 1940, 887 beginning October 1940, and 889 beginning January 1941; a smaller number of commodities was covered in earlier years. The price of each article is weighted by the approximate quantity marketed during a period approximating that covered by the index. Beginning with 1921, in computing indexes for commodity groups, articles falling under more than one of the classifications were included under each classification. For example, articles produced on the farm which reach the consumer practically unchanged in form, such as potatoes, milk, and eggs, were included among both farm products and foods. However, in computing the index for all commodities such articles were counted only once.

The retail-food-cost indexes presented in this section for periods beginning January 1935 are revised indexes based on the distribution of expenditures as shown by the 1934-36 study made by the Bureau of Labor Statistics of expenditures of wage earners' and lower salaried workers' families. Differences in changes in retail-food costs as shown by the revised, as compared with the unrevised, data are due largely to the relatively greater importance of citrus fruits and green vegetables and the lesser importance of cereals, potatoes, and apples in the revised index. The revised indexes are computed from prices of 54 foods. Aggregate costs of 54 foods in each of 51 cities, weighted to represent total purchases, have been combined for the United States with the use of population weights. In accordance with a recommendation of the Central Statistical Board, Bureau of the Budget, an average of the years 1935-39 is being used as a base in presenting these revised indexes of food costs. Indexes for all periods prior to January 1935 are converted from indexes computed for corresponding periods on the 1923-25 base.

The farm-price index is based on prices paid to producers for 34 major crops and 13 commercial truck crops. Average quotations for the period August 1909-July 1914 are used as a base, and each price series is weighted by the average annual marketings of farmers in the years 1924 to 1929.

In formulas for index numbers, it is customary to use p_0 and q_0, respectively, to designate the price and the quantity of a given commodity in the base year. For a second commodity, we should use p'_0 and q'_0, and so on. For the first commodity in the first year under consideration, we should use p_1 and q_1; for the second commodity, p'_1 and q'_1, etc., as shown in Fig. 36.

DESIGNATION OF PRICES AND QUANTITIES OF COMMODITIES

	Base Year or Period	Years under Consideration: First	Second	Third	Fourth	...	nth
Price of							
First commodity	p_0	p_1	p_2	p_3	p_4	...	p_n
Second commodity	p'_0	p'_1	p'_2	p'_3	p'_4	...	p'_n
Third commodity	p''_0	p''_1	p''_2	p''_3	p''_4	...	p''_n
Fourth commodity	p'''_0	p'''_1	p'''_2	p'''_3	p'''_4	...	p'''_n
...	...	...	...	...	...	...	...
nth commodity	$p^n{}_0$	$p^n{}_1$	$p^n{}_2$	$p^n{}_3$	$p^n{}_4$	...	$p^n{}_n$
Quantity of							
First commodity	q_0	q_1	q_2	q_3	q_4	...	q_n
Second commodity	q'_0	q'_1	q'_2	q'_3	q'_4	...	q'_n
Third commodity	q''_0	q''_1	q''_2	q''_3	q''_4	...	q''_n
Fourth commodity	q'''_0	q'''_1	q'''_2	q'''_3	q'''_4	...	q'''_n
...	...	...	...	...	...	...	...
nth commodity	$q^n{}_0$	$q^n{}_1$	$q^n{}_2$	$q^n{}_3$	$q^n{}_4$	...	$q^n{}_n$

Fig. 36

* *Statistical Abstract of the United States, 1942, p.* 371.

The simplest index is computed by dividing the sum of the prices (or quantities) of the commodities for the year or period under consideration by the sum of the prices (or quantities) of the same commodities in the base period:

SIMPLE AGGREGATIVE PRICE INDEX

$$\frac{\Sigma p_1}{\Sigma p_0}=\frac{p_1+p_1'+p_1''+\ldots+p_1^n}{p_0+p_0'+p_0''+\ldots+p_0^n}. \qquad \text{I}$$

In such computations, we run the risk that our findings will be dominated by the largest items, nullifying the effect of the smaller items.

For instance, in Table 34 we have one item which greatly exceeds the others in value. Though all items fluctuate, it is obvious that the index based on the totals is dependent largely on this first item. Here the indices are computed by dividing each total by the total for the base year, 1941.

TABLE 34

SIMPLE AGGREGATIVE PRICE INDEX

ITEMS	1941 p_0	1942 p_1	1943 p_2	1944 p_3
First	$6.31	$6.42	$6.48	$6.61
Second	0.27	0.29	0.30	0.32
Third	0.41	0.44	0.46	0.49
Fourth	0.11	0.11	0.12	0.14
Fifth	0.63	0.70	0.74	0.79
Totals	$7.73	$7.96	$8.10	$8.35
Indices	100.0	102.9	104.7	108.0

We may avoid the inequality caused by the preponderance of the cost of the first commodity in each instance by utilizing a different formula, the simple average of relatives price index:

SIMPLE AVERAGE OF RELATIVES PRICE INDEX

$$\frac{\Sigma\left(\frac{p_1}{p_0}\right)}{n}=\frac{\frac{p_1}{p_0}+\frac{p_1'}{p_0'}+\frac{p_1''}{p_0''}+\ldots+\frac{p_1^n}{p_0^n}}{n} \qquad \text{II}$$

Here our computations for the same figures would be:

TABLE 35

SIMPLE AVERAGE OF RELATIVES PRICE INDEX

IIEM	1941 Actual Price p_0	1941 Relative Price $\frac{p_0}{p_0}$	1942 Actual Price p_1	1942 Relative Price $\frac{p_1}{p_0}$	1943 Actual Price p_2	1943 Relative Price $\frac{p_2}{p_0}$	1944 Actual Price p_3	1944 Relative Price $\frac{p_3}{p_0}$
First	$6.31	100.0	$6.42	101.7	$6.48	102.6	$6.61	104.7
Second	0.27	100.0	0.29	107.4	0.30	111.1	0.32	118.5
Third	0.41	100.0	0.44	107.4	0.46	111.2	0.49	119.5
Fourth	0.11	100.0	0.11	100.0	0.12	109.1	0.14	127.3
Fifth	0.63	100.0	0.70	111.1	0.74	117.4	0.79	125.4
Totals		500.0		527.6		552.4		595.4
Indices		100.0		105.5		110.5		119.1

Instead of the "average", or arithmetic mean, we may employ either the median (p. 1193) or the geometric mean (p. 1198), if we so desire, with the same advantages and disadvantages already noted in Fig. 17 (p. 1202).

These figures would be satisfactory if fairly equal amounts of all five items had been sold. If the amounts differ markedly, we should employ a weighted average.

Using the quantities produced in the base year as the weights, we have the weighted aggregative, sometimes known as Laspeyres' formula:

WEIGHTED AGGREGATIVE:

$$\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} = \frac{p_1 q_0}{p_0 q_0} + \frac{p_1' q_0'}{p_0' q_0'} + \frac{p_1'' q_0''}{p_0'' q_0''} + \ldots + \frac{p_1^n q_0^n}{p_0^n q_0^n}. \qquad \text{III}$$

TABLE 36

WEIGHTED AGGREGATIVE—BASE-YEAR WEIGHTS

Item	Quantity	1941 Actual Price	1941 Price Times Quantity	1942 Actual Price	1942 Price Times Quantity	1943 Actual Price	1943 Price Times Quantity
	q_0	p_0	p_0q_0	p_1	p_1q_0	p_2	p_2q_0
First	25	$6.31	$157.95	$6.42	$160.50	$6.48	$162.00
Second	18	0.27	4.86	0.29	5.22	0.30	5.40
Third	47	0.41	19.27	0.44	20.68	0.46	21.42
Fourth	34	0.11	3.74	0.11	3.74	0.12	4.08
Fifth	7	0.63	4.41	0.70	4.90	0.74	5.18
Totals			$190.23		$195.04		$198.08
Indices			100.0		102.5		104.1

For some purposes, the weighted average of price relatives gives a clearer picture.

WEIGHTED AVERAGE OF PRICE RELATIVES

$$\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} = \frac{p_1}{p_0} \cdot p_0 q_0 + \frac{p_1'}{p_0'} \cdot p_0' q_0' + \frac{p_1''}{p_0''} \cdot p_0'' q_0'' + \ldots + \frac{p_1^n}{p_0^n} \cdot p_0^n q_0^n. \qquad \text{IV}$$

TABLE 37

WEIGHTED AVERAGE OF PRICE RELATIVES

Item	Quantity	1941 Actual Price	1941 Relative Price	1941 Price Times Quantity	1942 Actual Price	1942 Relative Price	1942 Relative Price Times Base Product	1943 Actual Price	1943 Relative Price	1943 Relative Price Times Base Product
	q_0	p_0	$\frac{p_0}{p_0}$	p_0q_0	p_1	$\frac{p_1}{p_0} \cdot p_0q_0$	$\frac{p_1}{p_0}$	p_2	$\frac{p_2}{p_0}$	$\frac{p_2}{p_0} \cdot p_0q_0$
First	25	$6.31	100.0	$157.95	$6.42	101.7	$160.63	$6.48	102.6	$162.06
Second	18	0.27	100.0	4.86	0.29	107.4	5.22	0.30	111.1	5.40
Third	47	0.41	100.0	19.27	0.44	107.4	20.70	0.46	112.2	21.62
Fourth	34	0.11	100.0	3.74	0.11	100.0	3.74	0.12	109.1	4.08
Fifth	7	0.63	100.0	4.41	0.70	111.1	4.90	0.74	117.4	5.18
Totals				$190.23			$195.19			$198.34
Indices				100.0			102.6			104.3

Since the quantities sold may differ greatly from year to year, it is desirable to take them into account in the weighting.

INDEX OF TOTAL VALUE

$$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} = \frac{p_1 q_1}{p_0 q_0} + \frac{p_1' q_1'}{p_0' q_0'} + \frac{p_1'' q_1''}{p_0'' q_0''} + \ldots + \frac{p_1^n q_1^n}{p_0^n q_0^n}. \qquad \text{V}$$

TABLE 38

INDEX OF TOTAL VALUE

Item	1941 Actual Price	1941 Quantity	1941 Price Times Quantity	1942 Actual Price	1942 Quantity	1942 Price Times Quantity	1943 Actual Price	1943 Quantity	1943 Price Times Quantity
	p_0	q_0	p_0q_0	p_1	q_1	p_1q_1	p_2	q_2	p_2q_2
First	$6.31	25	$157.95	$6.42	27	$173.34	$6.48	28	$181.44
Second	0.27	18	4.86	0.29	19	5.51	0.30	21	6.30
Third	0.41	47	19.27	0.44	49	21.56	0.46	50	23.00
Fourth	0.11	34	3.74	0.11	39	4.29	0.12	44	5.28
Fifth	0.63	7	4.41	0.70	8	5.60	0.74	8	5.92
Totals			$190.23			$210.30			$221.94
Indices			100.0			110.5			116.6

The index of total value gives flexibility, especially in times of a rapidly changing economy. Where weights are changed frequently to accommodate the structure of the index to the changing structure of the business, one cannot ascertain, for example, whether a change in the price index is due to a change in prices or to a change in weights. To overcome the danger of too great flexibility, the weights may be changed only occasionally, being held constant for a period of years.

All told, there are over 150 different formulas for the construction of index numbers now in actual use. Each of them presents some advantages, as noted. Unfortunately, each of them leads to a different result, so that one must be aware of the method by which the index is secured before one can give much credence to it. Probably any one of them is a better guide than a mere casual inspection and comparison of the data would give. If the same method is used throughout an investigation, there is at least a basis for comparison. The ideal formula, as conceived by Professor Irving Fisher, a pioneer and outstanding authority on index numbers, is:

Ideal Index Number Formula

$$\sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0}\times\frac{\Sigma p_1q_1}{\Sigma p_1q_0}}. \qquad \text{VI}$$

Other index formulas which are in very common use are:

Simple Average of Relatives, Physical Volume Index

$$\frac{\Sigma\left(\frac{q_1}{q_0}\right)}{n} \qquad \text{VII}$$

Weighted Aggregative Physical Volume Index

$$\frac{\Sigma p_0q_1}{\Sigma p_0q_0}=\frac{p_0q_1+p_0'q_1'+p_0''q_1''+\ldots+p_0^nq_1^n}{p_0q_0+p_0'q_0'+p_0''q_0''+\ldots+p_0^nq_0^n} \qquad \text{VIII}$$

Weighted Average of Relatives, Physical Volume Index

$$\frac{\frac{q_1}{q_0}\cdot p_0q_0+\frac{q_1'}{q_0'}\cdot p_0'q_0'+\frac{q_1''}{q_0''}\cdot p_0''q_0''+\ldots+\frac{q_1^n}{q_0^n}\cdot p_0^nq_0^n}{p_0q_0+p_0'q_0'+p_0''q_0''+\ldots+p_0^nq_0^n} \qquad \text{IX}$$

Chained Link Relative Index

$$I_i=\frac{\Sigma p_iq_0}{\Sigma p_{i-1}q_0}\cdot I_{i-1} \qquad \text{where } i=\text{current month} \qquad \text{X}$$

Because of the uncertainties and abnormal conditions during wartime, some of the standard index reports, like the production and trade indices formerly issued by the Federal Reserve Bank of New York, have been suspended for the duration.

FUNCTIONAL RELATIONSHIPS

At times when we are concerned with two or more variables, we are interested in discovering the extent to which the factors move together in the same direction or tend to move apart. If it can be shown that the values of one variable depend to a certain extent on the values of another variable or upon a combination of the values of several other variables, we say that there is a functional relationship between or among them. When such a functional relationship is treated statistically, we are using the method of correlation analysis.

This method is used whenever we wish to determine the possibility that there is a relationship, or association, between two items which have been noted in the compilation of the data.

Let us begin our consideration of the process with a simple experiment which can be repeated, with whatever variations are desired, as many times as one wishes.

Illustrative Example

In this experiment, the author took four coins of one denomination and six of another, shook them together, and noted the number of heads and tails in each fall. In order to compensate for the tendency of a particular coin to fall always head up because of the way the metal had worn, different coins were taken at intervals of from 15 to 35 throws. Sometimes pennies and nickels were used: at other times, dimes and quarters. The count was taken for each denomination separately and also for the total on each throw, as shown in Table 39.

TABLE 39

OCCURRENCE OF HEADS AND TAILS IN 150 THROWS

Number of Throws:	35	100	115	150
	Four-Coin Throw			
4 tails	2	10	10	12
1 head, 3 tails	10	13	18	28
2 heads, 2 tails	13	50	55	68
3 heads, 1 tail	8	20	25	33
4 heads	2	7	7	9
	Six-Coin Throw			
6 tails	1	2	2	3
1 head, 5 tails	3	12	12	15
2 heads, 4 tails	7	14	16	23
3 heads, 3 tails	18	29	36	54
4 heads, 2 tails	4	31	37	41
5 heads, 1 tail	1	11	11	12
6 heads	1	1	1	2

Number of Throws:	35	100	115	150
	Ten-Coin Throw			
10 tails	0	0	0	0
1 head, 9 tails	0	1	1	1
2 heads, 8 tails	2	5	5	7
3 heads, 7 tails	3	10	11	14
4 heads, 6 tails	9	15	17	26
5 heads, 5 tails	13	26	32	45
6 heads, 4 tails	5	26	30	35
7 heads, 3 tails	2	11	13	15
8 heads, 2 tails	1	5	5	6
9 heads, 1 tail	0	1	1	1
10 heads	0	0	0	0

Plotting the four-coin throw in Fig. 37, we note that, even for 35 cases, our curve tends to take the shape of the normal curve of distribution (p. 1188). As the number of cases is increased, the peak rises more perceptibly.

In the case of the six-coin throw, plotted in Fig. 38, the curve is somewhat skewed to the left after 35 throws, skews noticeably to the right after 100 and 115 throws, but, by the time 150 throws are reached, the peak is at the center of the diagram and the curve is approximating the symmetrical form which we should expect it eventually to assume. Because our distribution

range now covers seven cases, as against five in the previous figure, the peak does not rise quite so high.

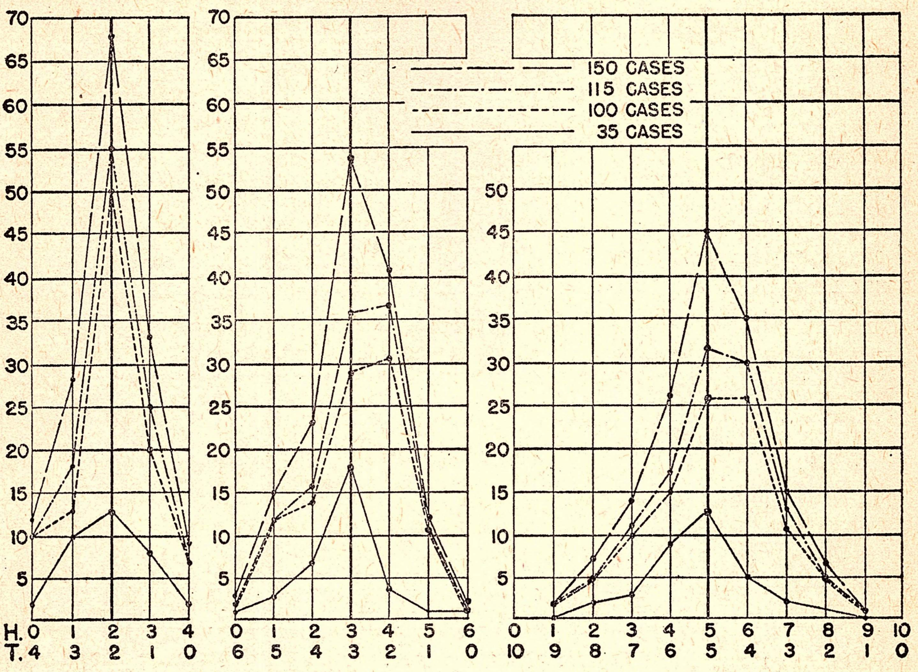


Fig. 37 Fig. 38 Fig. 39

When we consider the ten-coin throw, shown in Fig. 39, we find at all times a closer resemblance to the normal curve of distribution, the peak always at the center and the curve for each increased number of throws tending more and more to approach the normal form. The slight "hump" on the right, almost exactly balanced by a depression on the left, has decreased to such an extent that we may safely predict its eventual disappearance.

In Fig. 40, we have graphed on the same pair of axes the curves, after 150 throws, for the three different cases we have been considering. Here we note with greater ease the drop in the peak as the number of different combinations increases. In this figure, the ordinates line (frequencies) is always the same; the base-line is adjusted for each plotting so that the median case (even heads and tails) always falls on the same line. As we obtain successively a wider base on which to distribute the number of cases, the clustering of cases at the mid-point becomes less, although the tendency persists.

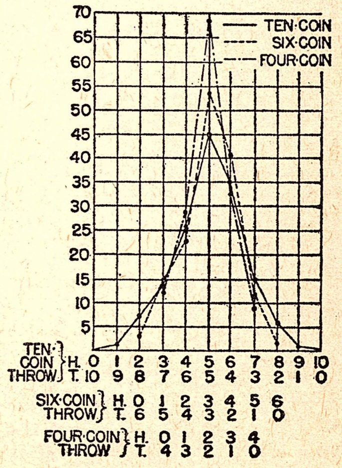


Fig. 40

From this, we see readily that, in any computation, we must have a sufficient number of cases ("samples") to enable us to draw accurate conclusions, but that an increase in the number of samples, after a certain point, does not change the picture very much. Judgment enters into this question, but there are some mathematical means of helping to determine the reliability of a sample and its sufficiency for the purpose at hand.

Scatter diagram

Preparation of a scatter diagram descriptive of the coin throws would show us something else about them which is of mathematical interest.

We plot data for one series on the X-axis and data for the other series on the Y-axis, drawing a diagonal across the graph and noting the relative distance of each item from this line.

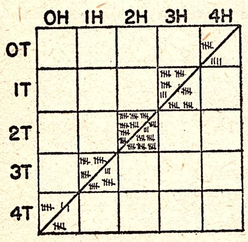


Fig. 41

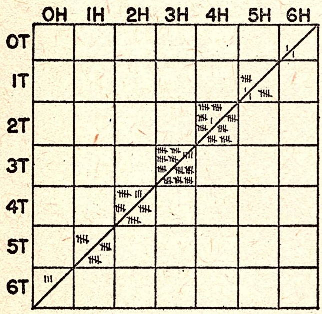


Fig. 42

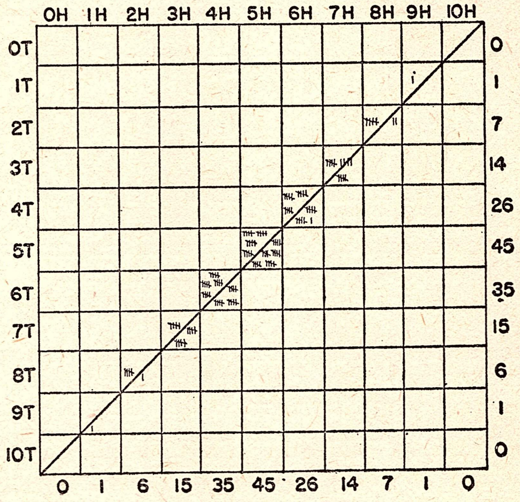


Fig. 43

In Fig. 41, we have a scatter diagram in which the various occurrences reported in Table 39 are indicated by tallies.* This diagram shows rather

*Scatter diagrams frequently make use of dots instead of tallies. By using tallies, however, one saves the labor of counting each separate dot in determining frequencies.

forcefully the cluster of throws about the point where the number of heads and the number of tails is equal, with a dropping off toward the ends of the scale. The diagonal line drawn across the diagram is the locus of that part of the diagram where all of the cases fall. Figs. 42 and 43 show the same tendencies for the six- and the ten-coin throws, except that, as we have previously noted, the lengthening of the scale causes a wider distribution and somewhat reduces the amount of cluster at the mid-point.

Thus far, we have been considering each kind of coin separately or (in the case of Figs. 39 and 43) the total occurrence of heads and tails without regard to the kind of coin. In Fig. 44, we have a scatter diagram in which the simultaneous occurrences are noted. While the frequency is still greatest along the diagonal and near the middle, the distribution no longer holds tight to the middle of the chart, but spreads in all directions, its intensity decreasing in proportion to its distance from the line of central tendency and from the mid-point of that line. Had we drawn our diagonal from the upper left-hand to the lower right-hand corner instead of from the lower left-hand to the upper right-hand corner, the same tendency would still persist.

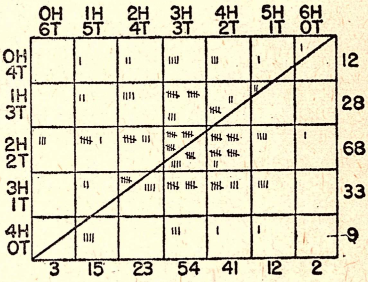


Fig. 44

This is the type of scatter diagram which we shall expect to secure when we compare two related series of data.

TEST YOUR ABILITY TO PREPARE SCATTER DIAGRAMS

1 Select 8 pennies and 8 nickels, keeping a record of the number of heads and tails for each on each successive throw. (a) Prepare graphs similar to Figs. 37 and 40, showing the situation after each 25 throws. (b) Prepare a scatter diagram showing the distribution of occurrences after your final throw. (You may carry this experiment as far as your patience persists. If you tally your results after every 25 or 50 throws, you will note the gradual accumulation of highest frequencies toward the center of your diagram. To compensate for the tendency of a particular coin always to fall in a given way, substitute other coins from time to time.)

The regression line

In Figs. 41 to 43, we noted the tendency of the items on the scatter diagram to follow a straight path across the diagram. This path is known as the line of regression. In these cases, it is a straight line; sometimes the observed facts will describe a curve, parabolic or otherwise.

STRAIGHT LINE OF REGRESSION

When the scatter diagram shows a straight-line relationship, we may calculate the line of regression by using a formula which bears a close resemblance to formula XXVI (p. 1224):

$$y_c = a + bx. \qquad \text{I}$$

The formula is the same as that for secular trend (p. 1224); the values of a and b are determined either by formulas XXVII and XXVIII or by taking:

$$a=\frac{\Sigma x^2\Sigma y-\Sigma x\Sigma xy}{n\Sigma x^2-(\Sigma x)^2} \quad * \qquad \text{II}$$

$$b=\frac{n\Sigma xy-\Sigma x\Sigma y}{n\Sigma x^2-(\Sigma x)^2} \quad * \qquad \text{III}$$

Table 40 shows a calculation of figures based on a comparison of the cost per pound of wool and the cost per pound of all-wool blankets,

COST OF WOOL PER POUND AND COST OF ALL-WOOL BLANKETS PER POUND

YEAR	BLANKETS (¢ PER LB.)		WOOL (¢ PER LB.)		PRODUCT
	y	y^2	x	x^2	xy
1935	120.1	14424.01	31.3	979.69	3759.13
1936	130.6	17056.36	40.9	1672.81	5341.54
1937	148.9	22171.21	43.3	1874.89	6447.37
1938	124.0	15376.00	29.4	864.36	3645.60
1939	126.5	16002.25	36.6	1339.56	4629.90
1940	146.3	21403.69	41.9	1755.61	6129.97
1941	155.3	24118.09	47.3	2237.29	7345.69
	$\Sigma y=951.7$	$\Sigma(y^2)=130551.61$	$\Sigma x=270.7$	$\Sigma(x^2)=10724.21$	$\Sigma(xy)=37299.20$

prepared to use these formulas. Taking the sums of the columns and substituting them in formulas II and III, we have

$$a=\frac{(10724.21)(951.7)-(270.7)(37299.20)}{7\ (10724.21)-(270.7)^2}=\frac{109337.217}{1790.98}=61.04882.$$

$$b=\frac{7\ (37299.20)-(270.7)(951.7)}{75069.47-73278.49}=\frac{3469.21}{1790.98}=1.93704.$$

Then formula I becomes $y_c=61.04882+1.93704x$.

By inserting in turn the various values of x from Table 40, we obtain the values which y would have in the same years if there were perfect agreement at all times between the wholesale cost of wool and the price of the finished blanket.

For 1935, for example, we should have

$$y_{c_{1935}}=61.04882+(1.93704)(31.3)=61.04882+60.629352=121.678172,$$

which is slightly higher than the actual price of 120.1 shown in the y column. The results of the remaining calculations are shown in Table 41, column 3.

TABLE 41

CALCULATION OF STANDARD ERROR OF ESTIMATE OF COST OF BLANKETS

YEAR	y	y_c	$y-y_c$	$(y-y_c)^2$
1935	120.1	121.7	−1.6	2.56
1936	130.6	140.3	−9.7	94.09
1937	148.9	144.9	4.0	16.00
1938	124.0	118.0	6.0	36.00
1939	126.5	131.9	−5.4	29.16
1940	146.3	142.2	4.1	16.81
1941	155.3	152.7	2.6	6.76
				$\Sigma(y-y_c)^2=201.38$

In expecting perfect agreement in such a regression line, we should be assuming that all other factors entering into the cost of blankets

*Where the original is taken at the middle of the time series, these reduce to $a=\frac{\Sigma y}{n}$ and $b=\frac{\Sigma xy}{\Sigma(x^2)}$.

remained constant. We realize without much reflection that such an assumption would be an absurdity. Since other factors (such as rising labor costs, increased costs of machinery and operations, consumer demand for blankets, etc.) may enter in, we expect that our estimate as found in the regression equation is somewhat in error. It is the extent of this error with which we are now concerned.

STANDARD ERROR OF ESTIMATE

By using formula I to compute a value for y (y_c) based on each of the given values of x, we have a basis by which to compare the deviations of the actual values from the estimated values of y_c. The process of determining the *standard error of estimate* is completed by adding the squares of these deviations, dividing by the number of cases, and extracting the square root:

$$S_y = \sqrt{\frac{\Sigma(y-y_c)^2}{n}}. \qquad \textbf{IVa}$$

Substituting the value of $(y-y_c)^2$ found in Table 41 in this formula, we have

$$S_y = \sqrt{\frac{201.38}{7}} = \sqrt{28.768} = \pm 5.3637.$$

Short method—A shorter way of arriving at the same result is given in formula IVb, which makes use of results already obtained in Table 40 and of the computations of the a and b values:

$$S_y = \sqrt{\frac{\Sigma(y^2) - (a\Sigma y + b\Sigma xy)}{n}}. \qquad \textbf{IVb}$$

Substituting these values, we have

$$S_y = \sqrt{\frac{131551.61 - [(61.04882)(951.7) + (1.93704)(37299.20)]}{7}}$$

$$= \sqrt{\frac{131551.61 - (58100.161994 + 72250.042368)}{7}}$$

$$= \sqrt{\frac{131551.61 - 130350.20462}{7}}$$

$$= \sqrt{\frac{201.4057}{7}} = \pm 5.3667,$$

which differs from our preceding calculation by an insignificant amount.

Corrected value—The corrected value of S_y (designated as $\bar{S}_y$), to overcome bias introduced by the calculations, is determined by taking into account the number of constants in the regression line (the constants being designated by m):

$$\bar{S}_y^2 = S_y^2\left(\frac{n}{n-m}\right) \qquad \textbf{V}$$

$$\bar{S}_y^2 = (5.36)^2\left(\frac{7}{7-2}\right) = 28.7296\left(\frac{7}{5}\right) = \frac{201.1072}{5} = 40.2214$$

Then $\bar{S}_y = \pm 6.34.$

TEST YOUR KNOWLEDGE OF REGRESSION LINES

2 The wholesale price of cotton, in cents per pound, was 11.9 in 1935, 12 in 1936, 11.3 in 1937, 8.7 in 1938, 9.1 in 1939, 10.1 in 1940, and 13.8 in 1941. For the same years, the wholesale price of cotton print cloth was: 1935, 4.9; 1936, 4.3; 1937, 4.8; 1938, 3.6; 1939, 3.6; 1940, 3.8; and 1941, 5.5. Calculate (a) the regression line; (b) the standard error of estimate of these figures, using both methods of computation.

3 From data of Table ζ, determine (a) the regression line and (b) the standard error of estimate when retail food prices are compared with wholesale.

TABLE ζ

PRICES, WHOLESALE, RETAIL, AND FARM—INDEXES, 1927-1941

(The paragraphs quoted on page 1238 explain the methods by which the figures in columns 2 and 4 were determined.)

Year	Wholesale Prices	Retail Food Prices	Farm Prices	House Furnishings	Clothing	Rent	Fuel, Electricity and Ice	Miscellaneous
1927	95.4	132.3	139	115.9	118.3	148.3	115.4	103.2
1928	96.7	130.8	149	113.1	116.5	144.8	113.4	103.8
1929	95.3	132.5	146	111.7	115.3	141.4	112.5	104.6
1930	86.4	126.0	126	108.9	112.7	137.5	111.4	105.1
1931	73.0	103.9	87	98.0	102.6	130.3	108.9	104.1
1932	64.8	86.5	65	85.4	90.8	116.9	103.4	101.7
1933	65.9	84.1	70	84.2	87.9	100.7	100.0	98.4
1934	74.9	93.7	90	92.8	96.1	94.4	101.4	97.9
1935	80.0	100.4	108	94.8	96.8	94.2	100.7	98.1
1936	80.8	101.3	114	96.3	97.6	96.4	100.2	98.7
1937	86.3	105.3	121	104.3	102.8	100.9	100.2	101.0
1938	78.6	97.8	95	103.3	102.2	104.1	99.9	101.5
1939	77.1	95.2	93	101.3	100.5	104.3	99.0	100.7
1940	78.6	96.6	98	100.5	101.7	104.6	99.7	101.1
1941	87.3	105.5	122	108.2	106.5	105.9	102.5	104.0

4 Again using Table ζ, calculate (a) the regression line and (b) the standard error of estimate when retail food prices are compared with farm prices.

Correlation

From the size of the standard error of estimate, we are able to determine how close the relationship between the series we are comparing really is. By dividing the standard error of estimate by the standard deviation, we make the factor of dispersion, $\frac{S_y}{\sigma_y}$, a constant. In a perfect relationship, since there are no deviations from the line of regression, $S_y=0$ and, as a result, $\frac{S_y}{\sigma_y}=0$. As the relationship becomes poorer, we approach 1 (or 100%) as a limit. Since we usually think of unity as indicating perfection, use of this value would require a complete reversal of our customary thinking. To avoid this, we subtract our ratio, $\frac{S_y}{\sigma_y}$, from 1, thus bringing our computation into line with what we should logically expect. Here a result of 0 indicates an imperfect relationship, while a result of 1 indicates perfect relationship. The nearer the result is to 1, the more nearly dependable is the relationship between the series.

LINEAR CORRELATION

In making use of the ratio between S_y and σ_y, we find that more accurate results are obtained by squaring each quantity before subtracting the ratio from 1. The result, called the *coefficient of correlation* and designated by r, is:

$$r = \sqrt{1 - \frac{(S_y)^2}{(\sigma_y)^2}}. \qquad \text{VI}$$

Squaring both sides gives us the *coefficient of determination*:

$$r^2 = 1 - \frac{(S_y)^2}{(\sigma_y)^2}. \qquad \text{VII}$$

Similarly, using k to denote the *coefficient of alienation*, we have:

$$k = \sqrt{1 - r^2} = \sqrt{1 - \left[1 - \frac{(S_y)^2}{(\sigma_y)^2}\right]} = \sqrt{\frac{(S_y)^2}{(\sigma_y)^2}} = \frac{S_y}{\sigma_y}, \qquad \text{VIII}$$

the other value which we discussed on page 1248.

When the b value for our regression line is negative, we take the negative square root as our value for r or k; when the b value is positive, we take the positive square root.

NON-LINEAR CORRELATION

In considering non-linear correlation, we use virtually the same formula as VI, but we distinguish our value as the *index of correlation*, expressing it as ρ (the Greek letter, rho):

$$\rho_{yx} = \sqrt{1 - \frac{(S_y)^2}{(\sigma_y)^2}}. \qquad \text{IX}$$

Note that it is necessary, when considering ρ, to indicate which variable is assumed to be dependent.

Just as we previously adjusted S_y (p. 1247), so also we may secure corrected estimates of r and ρ:

$$\bar{r}^2 = 1 - (1 - r^2)\left(\frac{n-1}{n-m}\right) \qquad \text{X}$$

and

$$\bar{\rho}^2_{yx} = 1 - (1 - \rho^2_{yx})\left(\frac{n-1}{n-m}\right). \qquad \text{XI}$$

In general, use of the curvilinear function rather than the straight line increases the per cent of variation, giving a better fit since the corrected standard error of estimate is smaller.

PRODUCT MOMENT METHOD

A simpler formula for obtaining r, known as the product moment method, is expressed by

$$r=\frac{p}{\sigma_x\sigma_y} \qquad \text{XII}$$

where (for ungrouped data)

$$p=\frac{\Sigma xy}{n}-\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)$$

$$\sigma_x=\sqrt{\frac{\Sigma(x^2)}{n}-\left(\frac{\Sigma x}{n}\right)^2}$$

$$\sigma_y=\sqrt{\frac{\Sigma(y^2)}{n}-\left(\frac{\Sigma y}{n}\right)^2}$$

and (for grouped data)

$$p=\frac{\Sigma f(d_x d_y)}{n}-\left(\frac{\Sigma fd_x}{n}\right)\left(\frac{\Sigma fd_y}{n}\right)$$

$$\sigma_x=\sqrt{\frac{\Sigma f(d_x)^2}{n}-\left(\frac{\Sigma fd_x}{n}\right)^2}$$

$$\sigma_y=\sqrt{\frac{\Sigma f(d_y)^2}{n}-\left(\frac{\Sigma fd_y}{n}\right)^2}.$$

When this method is used, the computation of p indicates the sign of r.

The typical change in y for a given change in x is found by

$$y=r\cdot\frac{\sigma_y}{\sigma_x}\cdot x \qquad \text{XIIIa}$$

while the similar formula to show the regression of x on y is

$$x=r\cdot\frac{\sigma_x}{\sigma_y}\cdot y\,. \qquad \text{XIIIb}$$

Use of this method eliminates the necessity of calculating the regression line and the standard error of estimate before determining r.

SPEARMAN FORMULAS

The rank method of measuring correlation consists of inspecting the two sets of data which are being compared and of assigning each item a number designating its order, or rank, when the items are ranged from highest to lowest. The differences between the two ranks (designated by $|D|$) are recorded, squared, and added. Then

$$\rho_S=1-\frac{6\Sigma(|D|^2)}{n(n^2-1)} \qquad \text{XIV}$$

In the case of two or more items having the same rank, we may use either the *bracket method*, in which all similar items are assigned the same number but the next lower item is assigned a position in the scale just as if they had been numbered differently, or the *mid-rank method*, in which two items which should each be numbered 4 (for example)

would be assigned the rank of 4.5, the next lower item being ranked 6, or three tied items (*e.g.*, 4,5,6) would all be ranked 5, etc.

In Table 42, where the item, "Number of customers, 99", appears three times, each appearance is ranked 4 since the next higher number, 104, was ranked 3; the next lower number, 97, however, ranks 7,

TABLE 42

RANK ORDER CORRELATION DATA FOR CUSTOMERS AND SALES—BRACKET METHOD

Department	Number of Customers	Amount of Sales	R_C	R_S	$\|D\|$	$\|D\|^2$
1	89	26.40	14	14	0	0
2	97	29.70	7	4	3	9
3	85	26.40	16	14	2	4
4	94	28.20	9	7	2	4
5	104	30.60	3	2	1	1
6	93	27.10	10	12	2	4
7	87	26.20	15	16	1	1
8	99	28.50	4	6	2	4
9	93	27.30	10	9	1	1
10	82	25.20	18	18	0	0
11	106	30.30	1	3	2	4
12	99	27.30	4	9	5	25
13	84	25.50	17	17	0	0
14	99	29.10	4	5	1	1
15	105	31.20	2	1	1	1
16	90	27.90	13	8	5	25
17	95	27.30	8	9	1	1
18	91	27.10	12	12	0	0
$n=18$					$\Sigma \|D\| = 29$	$\Sigma (\|D\|^2) = 85$

since there are *six* higher numbers in the list. By the mid-point method, shown in Table 43, the 99's are ranked as 5 (the median number in the 4-5-6 range). Similarly, in Table 42, 93 receives each

TABLE 43

RANK ORDER CORRELATION DATA FOR CUSTOMERS AND SALES—MID-RANK METHOD

Department	Number of Customers	Amount of Sales	R_C	R_S	$\|D\|$	$\|D\|^2$
1	89	26.40	14	14.5	0.5	0.25
2	97	29.70	7	4	3	9
3	85	26.40	16	14.5	1.5	2.25
4	94	28.20	9	7	2	4
5	104	30.60	3	2	1	1
6	93	27.10	10.5	12.5	2	4
7	87	26.20	15	16	1	1
8	99	28.50	5	6	1	1
9	93	27.30	10.5	10	0.5	0.25
10	82	25.20	18	18	0	0
11	106	30.30	1	3	2	4
12	99	27.30	5	10	5	25
13	84	25.50	17	17	0	0
14	99	29.10	4	5	1	1
15	105	31.20	2	1	1	1
16	90	27.90	13	8	5	25
17	95	27.30	8	10	2	2
18	91	27.10	12	12.5	0.5	0.25
$n=18$					$\Sigma \|D\| = 28$	$\Sigma(\|D\|^2) = 81$

time a rank of 10, while in Table 43, it receives a rank of 10.5. That there is little difference in the final outcome is shown by the calculations of the ρ value by the two methods.

Bracket Method:

$$\rho = 1 - \frac{6 \cdot 85}{18 \cdot 323} = 1 - \frac{510}{5814} = 1 - 0.087 = 0.913.$$

Mid-point Method:

$$\rho = 1 - \frac{6 \cdot 81}{18 \cdot 323} = 1 - \frac{486}{5814} = 1 - 0.083 = 0.917.$$

If the original data follow the normal distribution, we may compute r by taking

$$r=2\sin\left(\frac{\pi}{6}\rho\right). \qquad \text{XV}$$

Taking G to represent positive differences in rank, Spearman's *foot-rule* give us an approximation to the correlation:

$$R=1-\frac{6\Sigma G}{n^2-1}. \qquad \text{XVI}$$

By the product moment method, we find r directly:

$$r=\frac{\Sigma d_x d_y}{\sqrt{\Sigma(d_x)^2 \cdot \Sigma(d_y)^2}}. \qquad \text{XVII}$$

Computing the data as in Table 44, we arrive at

$$r=\frac{171.15}{\sqrt{(871)(-45.745)}}=0.857.$$

Here we first get the arithmetic mean of each rank, noting in the difference columns the amount by which each item deviates from the

TABLE 44

PRODUCT MOMENT DATA FOR CORRELATION OF CUSTOMERS AND SALES

Department	Number of Customers	Amount of Sales	d_x	d_y	$(d_x)^2$	$(d_y)^2$	d_xd_y
1	89	26.40	−5	−1.45	25	2.1025	7.25
2	97	29.70	3	1.85	9	3.4225	5.55
3	85	26.40	−9	−1.35	81	1.8225	12.15
4	94	28.20	0	0.35	0	0.1225	0
5	104	30.60	10	2.75	100	7.5625	27.50
6	93	27.10	−1	−0.75	1	0.5625	0.75
7	87	26.20	−7	−1.65	49	2.7225	11.55
8	99	28.50	5	0.65	25	0.4225	3.25
9	93	27.30	−1	−0.55	1	0.3025	0.55
10	82	25.20	−12	−2.65	144	7.0225	31.85
11	106	30.30	12	0.45	144	0.2025	5.40
12	99	27.30	5	−0.55	25	0.3025	−2.75
13	84	25.50	−10	−2.35	100	5.5225	23.50
14	99	29.10	5	1.25	25	1.5625	6.25
15	105	31.20	11	3.35	121	11.2225	36.85
16	90	27.90	−4	0.05	16	0.0025	−0.20
17	95	27.30	1	−0.55	1	0.3025	−0.55
18	91	27.10	−3	−0.75	4	0.5625	2.25
	$\Sigma_C=1692$	$\Sigma_S=501.30$			$\Sigma(d_x)^2=871$	$\Sigma(d_y)^2=45.745$	$\Sigma(d_xd_y)=171.15$
	$M_{A_C}=94$	$M_{A_S}=27.85$					

arithmetic mean. We square each of these deviations and also take the products of the comparable pairs, then summing them.

TEST YOUR KNOWLEDGE OF CORRELATON

5 Use the product moment to determine correlation between wholesale and retail prices, taking the data of Table ζ.

Multiple correlation

By multiple correlation, we may measure the relationship between a dependent variable and two or more independent variables. Here we use x_1 to indicate the dependent variable, and x_2, x_3, . . . , to indicate the independent variables.

LINEAR CORRELATION

The linear regression equation for two independent variables is

$$x_1 = a + b_{12\cdot3}x_2 + b_{13\cdot2}x_3, \; ^* \qquad \textbf{XVIII}$$

while for three independent variables it becomes

$$x_1 = a + b_{12\cdot34}x_2 + b_{13\cdot24}x_3 + b_{14\cdot23}x_4, \qquad \textbf{XIX}$$

and so on.

The normal equations for XIX are:

$$\Sigma x_1 = na + b_{12\cdot34}\Sigma x_2 + b_{13\cdot24}\Sigma x_3 + b_{14\cdot23}\Sigma x_4$$

$$\Sigma x_1x_2 = a\Sigma x_2 + b_{12\cdot34}\Sigma x_2^2 + b_{13\cdot24}\Sigma x_2x_3 + b_{14\cdot23}\Sigma x_2x_4$$

$$\Sigma x_1x_3 = a\Sigma x_3 + b_{12\cdot34}\Sigma x_2x_3 + b_{13\cdot24}\Sigma x_3^2 + b_{14\cdot23}\Sigma x_3x_4$$

$$\Sigma x_1x_4 = a\Sigma x_4 + b_{12\cdot34}\Sigma x_2x_4 + b_{13\cdot24}\Sigma x_3x_4 + b_{14\cdot23}\Sigma x_4^2 .$$

If we assume the origin of the equation to be at the point of averages, it follows that

$$p_{12} = b_{12\cdot34}\sigma_2^2 + b_{13\cdot24}p_{23} + b_{14\cdot23}p_{24}$$

$$p_{13} = b_{12\cdot34}p_{23} + b_{13\cdot24}\sigma_3^2 + b_{14\cdot23}p_{34}.$$

$$p_{14} = b_{12\cdot34} + b_{13\cdot24}p_{34} + b_{14\cdot23}\sigma_4^2 .$$

By solving these simultaneous equations, we secure the values of $b_{12\cdot34}$, $b_{13\cdot24}$, and $b_{14\cdot23}$.

The standard error of estimate is obtained from the formula,

$$S_{1\cdot234} = \sqrt{\frac{\Sigma(d^2)}{n}} \qquad \textbf{XXa}$$

or

$$S_{1\cdot234}^2 = \sigma_1^2 - b_{12\cdot34}p_{12} - b_{13\cdot24}p_{13} - b_{14\cdot23}p_{14} . \qquad \textbf{XXb}$$

In the same way, we obtain the coefficient of multiple correlation from

$$R_{1\cdot234} = \sqrt{1 - \frac{S^2_{1\cdot234}}{\sigma_1^2}} \qquad \textbf{XXIa}$$

or

$$R_{1\cdot234}^2 = \frac{b_{12\cdot34}p_{12} + b_{13\cdot24}p_{13} + b_{14\cdot23}p_{14}}{\sigma_1^2} . \qquad \textbf{XXIb}$$

The other coefficients (determination, non-determination, and alienation) are found as in simple correlation (p. 1249).

NON-LINEAR CORRELATION

When considering non-linear multiple correlation, we use fx_n to indicate any function of x_n. This function may take the form of any of the curves considered on pages 1223-1229.

$$x_1 = a + fx_2 + fx_3 + fx_4 + \; \ldots . \qquad \textbf{XXII}$$

---

*Note that $b_{12\cdot3}$ (read "b sub one two dot three", *not* "b twelve") indicates the number of units of change in the dependent variable (x) for a given unit of change in x_2, etc.

Partial correlation

Sometimes we wish to determine the separate effect or influence of each independent variable. To eliminate the effect of other independent variables, we use the *coefficient of partial correlation*:

$$r_{12\cdot 3}=\frac{r_{12}-r_{13}r_{23}}{(\sqrt{1-r_{13}^2})(\sqrt{1-r_{23}^2})} \qquad \text{XXIIIa}$$

$$r_{13\cdot 24}=\sqrt{b_{13\cdot 24}\cdot b_{31\cdot 24}} \qquad \text{XXIIIb}$$

$$r_{14\cdot 23}^2=1-\frac{1-R_{1\cdot 234}^2}{1-R_{1\cdot 23}^2} \text{ or } r_{14\cdot 23}=\sqrt{1-\frac{S_{1\cdot 234}^2}{S_{1\cdot 23}^2}} \qquad \text{XXIIIc}$$

These formulas should not be confused with the *coefficient of part correlation*, now not much used, which is expressed as

$$_{12}r_{34}=\frac{b_{12\cdot 34}^2\sigma_2^2}{b_{12\cdot 34}^2\sigma_2^2+\sigma_1^2(1-R_{1\cdot 234})}, \qquad \text{XXIV}$$

where the subscripts preceding the r indicate which variables have been excluded.

By means of *beta coefficients*, we may determine the relative importance of each of the independent variables. These are obtained by dividing each of the variables in the multiple regression equation,

$$x_1=a+b_{12\cdot 3}x_2+b_{13\cdot 2}x_3,$$

by its standard deviation, as

$$\frac{x_1}{\sigma_1}=a+b_{12\cdot 3}\cdot\frac{x_2}{\sigma_2}+b_{13\cdot 2}\cdot\frac{x_3}{\sigma_3}.$$

Then

$$\beta_{12\cdot 3}=b_{12\cdot 3}\cdot\frac{\sigma_2}{\sigma_1} \qquad \text{XXVa}$$

and

$$\beta_{13\cdot 2}=b_{13\cdot 2}\cdot\frac{\sigma_3}{\sigma_1} \qquad \text{XXVb}$$

INTERPRETATION OF CORRELATION

Measure	Formula for Measure	Formula for Standard Error of Measure	Formula for Probable Error of Measure
Coefficient of correlation	$r=\sqrt{1-\frac{(Sy)^2}{(\sigma y)^2}}$	$\sigma_r=\frac{1-r^2}{\sqrt{n}}$ $S_r=\frac{1-r^2}{\sqrt{n-2}}$	$PE_r=0.6745\cdot\frac{1-r^2}{\sqrt{n}}$
Index of correlation	$p=\sqrt{1-\frac{(Sy)^2}{(\sigma y)^2}}$	$\sigma_p=\frac{1-p^2}{\sqrt{n}}$	$PE_p=0.6745\cdot\frac{1-p^2}{\sqrt{n}}$
Coefficient of rank correlation	$p=1-\frac{6\Sigma(\lvert D\rvert^2)}{n(n^2-1)}$	$\sigma_p=\frac{1}{\sqrt{n-1}}$	$PE_p=0.6745\cdot\frac{1}{\sqrt{n-1}}$
Coefficient of multiple correlation	$R_{1\cdot 234}=\sqrt{1-\frac{S_{1\cdot 234}^2}{\sigma_1^2}}$	$\sigma R_{1\cdot 234}=\frac{1-R_{1\cdot 234}^2}{\sqrt{n}}$	$PE_R=0.6745\cdot\frac{1-R_{1\cdot 234}^2}{\sqrt{n}}$
Coefficient of partial correlation	r	$\sigma r_{12\cdot 34}=\frac{1-r_{12\cdot 34}^2}{\sqrt{n}}$	$PE_{r12\cdot 34}=0.6745\cdot\frac{1-r_{12\cdot 34}^2}{\sqrt{n}}$

Fig. 45

Tables and Formulas

TABLE XCIX

VALUES OF THE FUNCTION AND AREAS UNDER THE GAUSSIAN CURVE

Area under the Curve	t	Ordinate	Area under the Curve	t	Ordinate
0.0000	**0.00**	0.3989	0.4938	**2.50**	0.0175
.0199	**.05**	.3984	.4946	**2.55**	.0155
.0398	**.10**	.3970	.4953	**2.60**	.0136
.0596	**.15**	.3945	.4960	**2.65**	.0119
.0793	**.20**	.3910	.4965	**2.70**	.0104
.0987	**.25**	.3867	.4970	**2.75**	.0091
.1179	**.30**	.3814	.4974	**2.80**	.0079
.1368	**.35**	.3752	.4978	**2.85**	.0069
.1554	**.40**	.3683	.4981	**2.90**	.0060
.1736	**.45**	.3605	.4984	**2.95**	.0051
.1915	**.50**	.3521	.4987	**3.00**	.0044
.2088	**.55**	.3429	.4989	**3.05**	.0038
.2258	**.60**	.3332	.4990	**3.10**	.0033
.2422	**.65**	.3230	.4992	**3.15**	.0028
.2580	**.70**	.3123	.4993	**3.20**	.0024
.2734	**.75**	.3011	.4994	**3.25**	.0020
.2881	**.80**	.2897	.4995	**3.30**	.0017
.3023	**.85**	.2780	.4996	**3.35**	.0015
.3159	**.90**	.2661	.4997	**3.40**	.0012
.3289	**.95**	.2541	.4997	**3.45**	.0010
.3413	**1.00**	.2420	.4998	**3.50**	.0009
.3531	**1.05**	.2299	.4998	**3.55**	.0007
.3643	**1.10**	.2179	.4998	**3.60**	.0006
.3749	**1.15**	.2059	.4999	**3.65**	.0005
.3849	**1.20**	.1942	.4999	**3.70**	.0004
.3944	**1.25**	.1827	.4999	**3.75**	.0004
.4032	**1.30**	.1714	.4999	**3.80**	.0003
.4115	**1.35**	.1604	.4999	**3.85**	.0002
.4192	**1.40**	.1497	.5000	**3.90**	.0002
.4265	**1.45**	.1394	.5000	**3.95**	.0002
.4332	**1.50**	.1295	.5000	**4.00**	.0001
.4394	**1.55**	.1200	.5000	**4.05**	.0001
.4452	**1.60**	.1109	.5000	**4.10**	.0001
.4505	**1.65**	.1023	.5000	**4.15**	.0001
.4554	**1.70**	.0941	.5000	**4.20**	.0001
.4599	**1.75**	.0863	.5000	**4.25**	.0001
.4641	**1.80**	.0790	.5000	**4.30**	.0000
.4678	**1.85**	.0721	.5000	**4.35**	.0000
.4713	**1.90**	.0656	.5000	**4.40**	.0000
.4744	**1.95**	.0596	.5000	**4.45**	.0000
.4773	**2.00**	.0540	.5000	**4.50**	.0000
.4798	**2.05**	.0488	.5000	**4.55**	.0000
.4821	**2.10**	.0440	.5000	**4.60**	.0000
.4842	**2.15**	.0396	.5000	**4.65**	.0000
.4861	**2.20**	.0355	.5000	**4.70**	.0000
.4878	**2.25**	.0317	.5000	**4.75**	.0000
.4893	**2.30**	.0283	.5000	**4.80**	.0000
.4906	**2.35**	.0252	.5000	**4.85**	.0000
.4918	**2.40**	.0224	.5000	**4.90**	.0000
.4929	**2.45**	.0198	.5000	**4.95**	.0000

TABLE C
RECIPROCALS

n	$\frac{1}{n}$	n	$\frac{1}{n}$	n	$\frac{1}{n}$
1.0	1.00000	**4.0**	0.25000	**7.0**	0.14286
1.1	0.90909	**4.1**	.24390	**7.1**	.14085
1.2	.83333	**4.2**	.23810	**7.2**	.13889
1.3	.76923	**4.3**	.23256	**7.3**	.13699
1.4	.71429	**4.4**	.22727	**7.4**	.13514
1.5	.66667	**4.5**	.22222	**7.5**	.13333
1.6	.62500	**4.6**	.21379	**7.6**	.13158
1.7	.58823	**4.7**	.21277	**7.7**	.12987
1.8	.55556	**4.8**	.20833	**7.8**	.12821
1.9	.52632	**4.9**	.20408	**7.9**	.12658
2.0	.50000	**5.0**	.20000	**8.0**	.12500
2.1	.47619	**5.1**	.19608	**8.1**	.12346
2.2	.45455	**5.2**	.19231	**8.2**	.12195
2.3	.43478	**5.3**	.18868	**8.3**	.11976
2.4	.41667	**5.4**	.18519	**8.4**	.11905
2.5	.40000	**5.5**	.18182	**8.5**	.11765
2.6	.38462	**5.6**	.17857	**8.6**	.11628
2.7	.37037	**5.7**	.17544	**8.7**	.11494
2.8	.35714	**5.8**	.17241	**8.8**	.11364
2.9	.34483	**5.9**	.16949	**8.9**	.11236
3.0	.33333	**6.0**	.16667	**9.0**	.11111
3.1	.32258	**6.1**	.16393	**9.1**	.10989
3.2	.31250	**6.2**	.16129	**9.2**	.10870
3.3	.30303	**6.3**	.15873	**9.3**	.10753
3.4	.29412	**6.4**	.15625	**9.4**	.10638
3.5	.28571	**6.5**	.15385	**9.5**	.10526
3.6	.27778	**6.6**	.15152	**9.6**	.10417
3.7	.27027	**6.7**	.14925	**9.7**	.10309
3.8	.26316	**6.8**	.14706	**9.8**	.10204
3.9	.25641	**6.9**	.14493	**9.9**	.10101
				10.0	.10000

TABLE CI
FACTORS FOR ADJUSTING MONTHLY DATA FOR NUMBER OF CALENDAR DAYS PER MONTH

ORDINARY YEARS				LEAP YEARS		
Calendar Days	Ratio of Actual to Average Calendar Days	Reciprocal of Ratio	Month	Calendar Days	Ratio of Actual to Average Calendar Days	Reciprocal of Ratio
31	1.01918	98.11809	**January**	31	1.01639	98.35791
28	0.92055	108.63071	**February**	29	0.95082	105.17238
31	1.01918	98.11809	**March**	31	1.01639	98.35791
30	0.98630	101.38903	**April**	30	0.98367	101.66011
31	1.01918	98.11809	**May**	31	1.01639	98.35791
30	0.98630	101.38903	**June**	30	0.98367	101.66011
31	1.01918	98.11809	**July**	31	1.01639	98.35791
31	1.01918	98.11809	**August**	31	1.01639	98.35791
30	0.98630	101.38903	**September**	30	0.98367	101.66011
31	1.01918	98.11809	**October**	31	1.01639	98.35791
30	0.98630	101.38903	**November**	30	0.98367	101.66011
31	1.01918	98.11809	**December**	31	1.01639	98.35791
365				366		

TABLE CII

PROBABILITY OF THE OCCURRENCE OF STATISTICAL DEVIATIONS

When the number of standard errors is as given in the second column, the probability that a deviation as great as or greater than the designated number of standard errors will occur is shown in the first column and the likelihood that so great a deviation will not occur is shown in the third column.

Probability of Occurrence	Number of Standard Errors	Odds Against Occurrence	Probability of Occurrence	Number of Standard Errors	Odds Against Occurrence
50.00%	**0.67449**	1.00 : 1	1.24%	**2.5**	79.52 : 1
48.39	**0.7**	1.07 : 1	0.932	**2.6**	106.3 : 1
42.37	**0.8**	1.36 : 1	0.693	**2.7**	143.2 : 1
36.81	**0.9**	1.72 : 1	0.511	**2.8**	194.7 : 1
31.73	**1.0**	2.15 : 1	0.373	**2.9**	267.0 : 1
27.13	**1.1**	2.69 : 1	0.270	**3.0**	369.4 : 1
23.01	**1.2**	3.35 : 1	0.194	**3.1**	515.7 : 1
19.36	**1.3**	4.17 : 1	0.137	**3.2**	726.7 : 1
16.15	**1.4**	5.19 : 1	0.0967	**3.3**	1033 : 1
13.36	**1.5**	6.48 : 1	0.0674	**3.4**	1483 : 1
10.96	**1.6**	8.12 : 1	0.0465	**3.5**	2149 : 1
8.91	**1.7**	10.22 : 1	0.0318	**3.6**	3142 : 1
7.19	**1.8**	12.92 : 1	0.0216	**3.7**	4637 : 1
5.74	**1.9**	16.41 : 1	0.0145	**3.8**	6915 : 1
4.55	**2.0**	20.98 : 1	0.00962	**3.9**	10390 : 1
3.57	**2.1**	26.99 : 1	0.00634	**4.0**	15770 : 1
2.78	**2.2**	34.96 : 1	0.0000573	**5.0**	1744000 : 1
2.14	**2.3**	45.62 : 1	0.00000020	**6.0**	500000000 : 1
1.64	**2.4**	60.00 : 1	0.00000000026	**7.0**	400000000000 : 1

TABLE CIII

t TABLE FOR SMALL SAMPLES

$n-1$*	50%	95%	99%	$n-1$*	50%	95%	99%
1	1.000	12.706	63.657	**16**	0.690	2.120	2.921
2	0.816	4.303	9.925	**17**	0.689	2.110	2.898
3	0.765	3.182	5.841	**18**	0.688	2.101	2.878
4	0.741	2.776	4.604	**19**	0.688	2.093	2.861
5	0.727	2.571	4.032	**20**	0.687	2.086	2.845
6	0.718	2.447	3.707	**21**	0.686	2.080	2.831
7	0.711	2.365	3.499	**22**	0.686	2.074	2.819
8	0.706	2.306	3.355	**23**	0.685	2.069	2.807
9	0.703	2.262	3.250	**24**	0.685	2.064	2.797
10	0.700	2.228	3.169	**25**	0.684	2.060	2.787
11	0.697	2.201	3.106	**26**	0.684	2.056	2.779
12	0.695	2.179	3.055	**27**	0.684	2.052	2.771
13	0.694	2.160	3.012	**28**	0.683	2.048	2.763
14	0.692	2.145	2.977	**29**	0.683	2.045	2.756
15	0.691	2.131	2.947	**30**	0.683	2.042	2.750

*By taking the bold numerals to read "$n-2$", this table may be used also for the standard error of the coefficient of correlation in samples involving fewer than 30 cases.

TABLE CIV

VALUES OF $\chi_1 \equiv \dfrac{0.6745}{\sqrt{n}}$

n	χ_1	n	χ_1	n	χ_1	n	χ_1	n	χ_1
1	0.6745	51	0.0944	101	0.0671	155	0.0542	510	0.0299
2	.4769	52	.0935	102	.0668	160	.0533	520	.0296
3	.3894	53	.0926	103	.0665	165	.0525	530	.0293
4	.3372	54	.0918	104	.0661	170	.0517	540	.0290
5	.3016	55	.0909	105	.0658	175	.0510	550	.0288
6	.2754	56	.0901	106	.0655	180	.0503	560	.0285
7	.2549	57	.0893	107	.0652	185	.0496	570	.0283
8	.2385	58	.0886	108	.0649	190	.0489	580	.0280
9	.2248	59	.0878	109	.0646	195	.0483	590	.0278
10	.2133	60	.0871	110	.0643	200	.0477	600	.0275
11	.2034	61	.0864	111	.0640	205	.0471	610	.0273
12	.1947	62	.0857	112	.0637	210	.0465	620	.0271
13	.1871	63	.0850	113	.0635	215	.0460	630	.0269
14	.1803	64	.0843	114	.0632	220	.0455	640	.0267
15	.1742	65	.0837	115	.0629	225	.0450	650	.0265
16	.1686	66	.0830	116	.0626	230	.0445	660	.0263
17	.1636	67	.0824	117	.0624	235	.0440	670	.0261
18	.1590	68	.0818	118	.0621	240	.0435	680	.0259
19	.1547	69	.0812	119	.0618	245	.0431	690	.0257
20	.1508	70	.0806	120	.0616	250	.0427	700	.0255
21	.1472	71	.0800	121	.0613	255	.0422	710	.0253
22	.1438	72	.0795	122	.0611	260	.0418	720	.0251
23	.1406	73	.0789	123	.0608	265	.0414	730	.0250
24	.1377	74	.0784	124	.0606	270	.0410	740	.0248
25	.1349	75	.0779	125	.0603	275	.0407	750	.0246
26	.1323	76	.0774	126	.0601	280	.0403	760	.0245
27	.1298	77	.0769	127	.0599	285	.0400	770	.0243
28	.1275	78	.0764	128	.0596	290	.0396	780	.0242
29	.1252	79	.0759	129	.0594	295	.0393	790	.0240
30	.1231	80	.0754	130	.0592	300	.0389	800	.0238
31	.1211	81	.0749	131	.0589	310	.0383	810	.0237
32	.1192	82	.0745	132	.0587	320	.0377	820	.0236
33	.1174	83	.0740	133	.0585	330	.0371	830	.0234
34	.1157	84	.0736	134	.0583	340	.0366	840	.0233
35	.1140	85	.0732	135	.0581	350	.0361	850	.0231
36	.1124	86	.0727	136	.0578	360	.0355	860	.0230
37	.1109	87	.0723	137	.0576	370	.0351	870	.0229
38	.1094	88	.0719	138	.0574	380	.0346	880	.0227
39	.1080	89	.0715	139	.0572	390	.0342	890	.0226
40	.1066	90	.0711	140	.0570	400	.0337	900	.0225
41	.1053	91	.0707	141	.0568	410	.0333	910	.0224
42	.1041	92	.0703	142	.0566	420	.0329	920	.0222
43	.1029	93	.0699	143	.0564	430	.0325	930	.0221
44	.1017	94	.0696	144	.0562	440	.0322	940	.0220
45	.1005	95	.0692	145	.0560	450	.0318	950	.0219
46	.0994	96	.0688	146	.0558	460	.0314	960	.0218
47	.0984	97	.0685	147	.0556	470	.0311	970	.0217
48	.0974	98	.0681	148	.0554	480	.0308	980	.0215
49	.0964	99	.0678	149	.0553	490	.0305	990	.0214
50	.0954	100	.0674	150	.0551	500	.0302	1000	.0213

TABLE CV

FACTORS FOR COMPUTING PROBABLE ERROR

$\frac{0.6745}{\sqrt{n-1}}$	n	$\frac{0.6745}{\sqrt{n(n-1)}}$	$\frac{0.6745}{\sqrt{n-1}}$	n	$\frac{0.6745}{\sqrt{n(n-1)}}$
—	**1**	—	0.0954	**51**	0.0134
0.6745	**2**	0.4769	.0945	**52**	.0131
.4769	**3**	.2754	.0935	**53**	.0129
.3894	**4**	.1947	.0927	**54**	.0126
.3372	**5**	.1508	.0918	**55**	.0124
.3016	**6**	.1231	.0910	**56**	.0122
.2754	**7**	.1041	.0901	**57**	.0119
.2549	**8**	.0901	.0893	**58**	.0117
.2385	**9**	.0795	.0886	**59**	.0115
.2248	**10**	.0711	.0878	**60**	.0113
.2133	**11**	.0643	.0871	**61**	.0112
.2034	**12**	.0587	.0864	**62**	.0110
.1947	**13**	.0504	.0857	**63**	.0108
.1871	**14**	.0500	.0850	**64**	.0106
.1803	**15**	.0465	.0843	**65**	.0105
.1742	**16**	.0435	.0837	**66**	.0103
.1686	**17**	.0409	.0830	**67**	.0101
.1636	**18**	.0386	.0824	**68**	.0100
.1590	**19**	,9365	.0818	**69**	.0099
.1547	**20**	.0346	.0812	**70**	.0097
.1508	**21**	.0329	.0806	**71**	.0096
.1472	**22**	.0314	.0801	**72**	.0094
.1438	**23**	.0300	.0795	**73**	.0093
.1406	**24**	.0287	.0789	**74**	.0092
.1377	**25**	.0275	.0784	**75**	.0091
.1349	**26**	.0265	.0779	**76**	.0089
.1323	**27**	.0255	.0773	**77**	.0088
.1298	**28**	.0245	.0769	**78**	.0087
.1275	**29**	.0237	.0764	**79**	.0086
.1252	**30**	.0229	.0759	**80**	.0085
.1231	**31**	.0221	.0754	**81**	.0084
.1211	**32**	.0214	.0749	**82**	.0083
.1192	**33**	.0208	.0745	**83**	.0082
.1174	**34**	.0201	.0740	**84**	.0081
.1157	**35**	.0196	.0736	**85**	.0080
.1140	**36**	.0190	.0732	**86**	.0079
.1124	**37**	.0185	.0727	**87**	.0078
.1109	**38**	.0180	.0723	**88**	.0077
.1094	**39**	.0175	.0719	**89**	.0076
.1080	**40**	.0171	.0715	**90**	.0075
.1066	**41**	.0167	.0711	**91**	.0075
.1053	**42**	.0163	.0707	**92**	.0074
.1041	**43**	.0159	.0703	**93**	.0073
.1029	**44**	.0155	.0699	**94**	.0072
.1017	**45**	.0152	.0696	**95**	.0071
.1005	**46**	.0148	.0692	**96**	.0071
.0994	**47**	.0145	.0688	**97**	.0070
.0984	**48**	.0142	.0685	**98**	.0069
.0974	**49**	.0139	.0681	**99**	.0069
.0964	**50**	.0136	.0678	**100**	.0068

TABLE CVI
FACTORS FOR COMPUTING PROBABLE ERROR OF MEDIANS

$\frac{0.8453}{n(\sqrt{n}-1)}$	n	$\frac{0.8453}{\sqrt{n(n-1)}}$	$\frac{0.8453}{n(\sqrt{n}-1)}$	n	$\frac{0.8453}{\sqrt{n(n-1)}}$
—	1	—	0.0023	51	0.0167
0.4227	2	0.5978	.0023	52	.0164
.1993	3	.3451	.0022	53	.0161
.1220	4	.2440	.0022	54	.0158
.0845	5	.1890	.0021	55	.0155
.0630	6	.1543	.0020	56	.0152
.0493	7	.1304	.0020	57	.0150
.0399	8	.1130	.0019	58	.0147
.0332	9	.0996	.0019	59	.0145
.0282	10	.0891	.0018	60	.0142
.0243	11	.0806	.0018	61	.0140
.0212	12	.0736	.0018	62	.0138
.0188	13	.0677	.0017	63	.0135
.0167	14	.0627	.0017	64	.0133
.0151	15	.0583	.0016	65	.0131
.0136	16	.0546	.0016	66	.0129
.0124	17	.0513	.0016	67	.0127
.0114	18	.0483	.0015	68	.0125
.0105	19	.0457	.0015	69	.0123
.0097	20	.0434	.0015	70	.0122
.0090	21	.0412	.0014	71	.0120
.0084	22	.0393	.0014	72	.0118
.0078	23	.0376	.0014	73	.0117
.0073	24	.0360	.0013	74	.0115
.0069	25	.0345	.0013	75	.0113
.0065	26	.0332	.0013	76	.0112
.0061	27	.0319	.0013	77	.0111
.0058	28	.0307	.0012	78	.0109
.0055	29	.0297	.0012	79	.0108
.0052	30	.0287	.0012	80	.0106
.0050	31	.0277	.0012	81	.0105
.0047	32	.0268	.0012	82	.0104
.0045	33	.0260	.0011	83	.0103
.0043	34	.0252	.0011	84	.0101
.0041	35	.0245	.0011	85	.0100
.0040	36	.0238	.0011	86	.0099
.0038	37	.0232	.0011	87	.0098
.0037	38	.0225	.0010	88	.0097
.0035	39	.0220	.0010	89	.0096
.0034	40	.0214	.0010	90	.0094
.0033	41	.0209	.0010	91	.0093
.0031	42	.0204	.0010	92	.0092
.0030	43	.0199	.0010	93	.0091
.0029	44	.0194	.0009	94	.0090
.0028	45	.0190	.0009	95	.0089
.0027	46	.0186	.0009	96	.0089
.0027	47	.0182	.0009	97	.0088
.0026	48	.0178	.0009	98	.0087
.0025	49	.0174	.0009	99	.0086
.0024	50	.0171	.0008	100	.0085

TABLE CVII
PROBABLE ERROR OF THE PRODUCT MOMENT COEFFICIENT

n	.0	.2	.4	.6	.8
20	0.1508	0.1448	0.1267	0.0965	0.0543
40	.1066	.1024	.0896	.0683	.0384
60	.0871	.0836	.0731	.0557	.0313
80	.0754	.0724	.0633	.0483	.0271
100	.0674	.0648	.0567	.0432	.0243
120	.0616	.0591	.0517	.0394	.0222
140	.0570	.0547	.0479	.0365	.0205
160	.0533	.0512	.0448	.0341	.0192
180	.0503	.0483	.0422	.0322	.0181
200	.0477	.0458	.0401	.0305	.0172
220	.0455	.0437	.0382	.0291	.0164
240	.0435	.0418	.0366	.0279	.0157
260	.0418	.0402	.0351	.0268	.0151
280	.0403	.0387	.0339	.0258	.0145
300	.0389	.0374	.0327	.0249	.0140
320	.0377	.0362	.0317	.0241	.0136
340	.0366	.0351	.0307	.0234	.0132
360	.0355	.0341	.0299	.0228	.0128
380	.0346	.0332	.0291	.0221	.0125
400	.0337	.0324	.0283	.0216	.0121
420	.0329	.0316	.0276	.0211	.0118
440	.0322	.0309	.0270	.0206	.0116
460	.0314	.0302	.0264	.0201	.0113
480	.0308	.0296	.0259	.0197	.0111
500	.0302	.0290	.0253	.0193	.0109
520	.0296	.0284	.0248	.0189	.0106
540	.0290	.0279	.0244	.0186	.0105
560	.0285	.0274	.0239	.0182	.0103
580	.0280	.0269	.0235	.0179	.0101
600	.0275	.0264	.0231	.0176	.0099
620	.0271	.0260	.0228	.0173	.0098
640	.0267	.0256	.0224	.0171	.0096
660	.0263	.0252	.0221	.0168	.0095
680	.0259	.0248	.0217	.0166	.0093
700	.0255	.0245	.0214	.0163	.0092
720	.0251	.0241	.0211	.0161	.0090
740	.0248	.0238	.0208	.0159	.0089
760	.0245	.0235	.0206	.0157	.0088
780	.0242	.0232	.0203	.0155	.0087
800	.0238	.0229	.0200	.0153	.0086
820	.0236	.0226	.0198	.0151	.0085
840	.0233	.0223	.0195	.0149	.0084
860	.0230	.0221	.1093	.0147	.0083
880	.0227	.0218	.0191	.0146	.0082
900	.0225	.0216	.0189	.0144	.0081
920	.0222	.0213	.0187	.0142	.0080
940	.0220	.0211	.0185	.0141	.0079
960	.0218	.0209	.0183	.0139	.0078
980	.0215	.0207	.0181	.0138	.0078
1000	.0213	.0205	.0179	.0137	.0077

TABLE CVIII
CORRELATION VALUES

r	$\sqrt{1-r}$	r	$\sqrt{1-r}$	r	$\sqrt{1-r}$	r	$\sqrt{1-r}$
0.000	1.0000	0.250	0.8660	0.500	0.7071	0.750	0.5000
.005	.9975	.255	.8631	.505	.7036	.755	.4950
.010	.9950	.260	.8602	.510	.7000	.760	.4899
.015	.9925	.265	.8573	.515	.6964	.765	.4848
.020	.9900	.270	.8544	.520	.6928	.770	.4796
.025	.9874	.275	.8515	.525	.6892	.775	.4743
.030	.9849	.280	.8485	.530	.6856	.780	.4690
.035	.9823	.285	.8456	.535	.6819	.785	.4637
.040	.9798	.290	.8426	.540	.6782	.790	.4583
.045	.9772	.295	.8396	.545	.6745	.795	.4528
.050	.9747	.300	.8367	.550	.6708	.800	.4472
.055	.9721	.305	.8337	.555	.6671	.805	.4416
.060	.9695	.310	.8307	.560	.6633	.810	.4359
.065	.9670	.315	.8277	.565	.6596	.815	.4301
.070	.9643	.320	.8246	.570	.6557	.820	.4243
.075	.9618	.325	.8216	.575	.6519	.825	.4183
.080	.9592	.330	.8185	.580	.6481	.830	.4123
.085	.9566	.335	.8155	.585	.6442	.835	.4062
.090	.9539	.340	.8124	.590	.6403	.840	.4000
.095	.9513	.345	.8093	.595	.6364	.845	.3937
.100	.9487	.350	.8062	.600	.6325	.850	.3873
.105	.9461	.355	.8031	.605	.6285	.855	.3808
.110	.9434	.360	.8000	.610	.6245	.860	.3742
.115	.9407	.365	.7969	.615	.6205	.865	.3674
.120	.9381	.370	.7937	.620	.6164	.870	.3606
.125	.9354	.375	.7906	.625	.6124	.875	.3536
.130	.9327	.380	.7874	.630	.6083	.880	.3461
.135	.9301	.385	.7842	.635	.6042	.885	.3391
.140	.9274	.390	.7810	.640	.6000	.890	.3317
.145	.9247	.395	.7778	.645	.5958	.895	.3240
.150	.9220	.400	.7746	.650	.5916	.900	.3162
.155	.9192	.405	.7714	.655	.5874	.905	.3082
.160	.9165	.410	.7681	.660	.5831	.910	.3000
.165	.9138	.415	.7649	.665	.5788	.915	.2916
.170	.9110	.420	.7616	.670	.5745	.920	.2828
.175	.9083	.425	.7583	.675	.5701	.925	.2739
.180	.9055	.430	.7550	.680	.5657	.930	.2646
.185	.9028	.435	.7517	.685	.5613	.935	.2550
.190	.9000	.440	.7483	.690	.5568	.940	.2450
.195	.8972	.445	.7450	.695	.5523	.945	.2345
.200	.8944	.450	.7416	.700	.5477	.950	.2236
.205	.8916	.455	.7382	.705	.5431	.955	.2121
.210	.8888	.460	.7349	.710	.5385	.960	.2000
.215	.8860	.465	.7314	.715	.5339	.965	.1871
.220	.8832	.470	.7280	.720	.5292	.970	.1732
.225	.8803	.475	.7246	.725	.5244	.975	.1581
.230	.8775	.480	.7211	.730	.5196	.980	.1414
.235	.8746	.485	.7176	.735	.5148	.985	.1225
.240	.8718	.490	.7141	.740	.5099	.990	.1000
.245	.8689	.495	.7106	.745	.5050	.995	.0707

TABLE CIX
COMPUTATIONAL VALUES

$\frac{1}{\sqrt{n}}$	n	$\frac{1}{\sqrt{n(n-1)}}$	$\frac{1}{\sqrt{n}}$	n	$\frac{1}{\sqrt{n(n-1)}}$
1	1	. .	0.14003	51	0.01980
0.70711	2	0.70711	.13868	52	.01942
.57735	3	.40825	.13736	53	.01905
.50000	4	.28868	.13608	54	.01869
.44721	5	.22361	.13484	55	.01835
.40825	6	.18257	.13363	56	0.1802
.37796	7	.15430	.13245	57	0.1770
.35355	8	.13363	.13131	58	.01739
.33333	9	.11785	.13019	59	.01710
.31623	10	.10541	.12910	60	.01681
.30151	11	.09535	.12804	61	.01653
.28868	12	.08704	.12700	62	.01626
.27735	13	.08006	.12599	63	.01600
.26726	14	.07413	.12500	64	.01575
.25820	15	.06901	.12404	65	.01550
.25000	16	.06455	.12309	66	.01527
.24254	17	.06063	.12217	67	.01504
.23570	18	.05717	.12127	68	.01482
.22942	19	.05407	.12039	69	.01460
.22361	20	.05130	.11952	70	.01439
.21822	21	.04880	.11868	71	.01419
.21320	22	.04652	.11785	72	.01399
.20851	23	.04446	.11704	73	.01379
.20412	24	.04256	.11625	74	.01361
.20000	25	.04083	.11547	75	.01342
.19612	26	.03922	.11471	76	.01325
.19245	27	.03774	.11396	77	.01307
.18898	28	.03637	.11323	78	.01290
.18570	29	.03509	.11251	79	.01274
.18257	30	.03390	.11180	80	.01258
.17961	31	.03279	.11111	81	.01242
.17678	32	.03175	.11043	82	.01227
.17408	33	.03077	.10976	83	.01212
.17150	34	.02985	.10911	84	.01198
.16903	35	.02899	.10847	85	.01184
.16667	36	.02817	.10783	86	.01170
.16440	37	.02740	.10721	87	.01156
.16222	38	.02667	.10660	88	.01143
.16013	39	.02598	.10600	89	.01130
.15811	40	.02532	.10541	90	.01117
.15617	41	.02469	.10483	91	.01105
.15430	42	.02409	.10426	92	.01093
.15250	43	.02353	.10370	93	.01081
.15076	44	.02299	.10314	94	.01070
.14907	45	.02247	.10260	95	.01058
.14744	46	.02198	.10206	96	.01047
.14587	47	.02151	.10154	97	.01036
.14434	48	.02105	.10102	98	.01026
.14286	49	.02062	.10050	99	.01015
.14142	50	.02020	.10000	100	.01005

TABLE CIX (Continued)
COMPUTATIONAL VALUES

$\sqrt{10n}$	n	$\sqrt[3]{10n}$	$\sqrt{10n}$	n	$\sqrt[3]{10n}$
3.16228	**1**	2.15444	22.58318	**51**	7.98957
4.47214	**2**	2.71442	22.80351	**52**	8.04145
5.47723	**3**	3.10723	23.02173	**53**	8.09267
6.32456	**4**	3.41995	23.23790	**54**	8.14325
7.07107	**5**	3.68403	23.45208	**55**	8.19321
7.74597	**6**	3.91487	23.66432	**56**	8.24257
8.36660	**7**	4.12129	23.87467	**57**	8.29134
8.94427	**8**	4.30887	24.08319	**58**	8.33955
9.48683	**9**	4.48141	24.28992	**59**	8.38721
10.00000	**10**	2.15444	24.49490	**60**	8.43433
10.48809	**11**	4.79142	24.69818	**61**	8.48093
10.95445	**12**	4.93242	24.89980	**62**	8.52702
11.40175	**13**	5.06580	25.09980	**63**	8.57262
11.83216	**14**	5.19249	25.29822	**64**	8.61774
12.24745	**15**	5.31329	25.49510	**65**	8.66239
12.64911	**16**	5.42884	25.69047	**66**	8.70659
13.03840	**17**	5.53966	25.88436	**67**	8.75034
13.41641	**18**	5.64622	26.07681	**68**	8.79366
13.78405	**19**	5.74890	26.26785	**69**	8.83656
14.14214	**20**	5.84804	26.45751	**70**	8.87904
14.49138	**21**	5.94392	26.64583	**71**	8,92112
14.83240	**22**	6.03681	26.83282	**72**	8.96281
15.16575	**23**	6.12693	27.01851	**73**	9.00411
15.49193	**24**	6.21447	27.20294	**74**	9.04504
15.81139	**25**	6.29961	27.38613	**75**	9.08560
16.12452	**26**	6.38250	27.56810	**76**	9.12581
16.43168	**27**	6.46330	27.74887	**77**	9.16566
16.73320	**28**	6.54213	27.92848	**78**	9.20516
17.02939	**29**	6.61911	28.10694	**79**	9.24434
17.32051	**30**	6.69433	28.28427	**80**	9.28318
17.60682	**31**	6.76790	28.46050	**81**	9.32170
17.88854	**32**	6.83990	28.63564	**82**	9.35990
18.16590	**33**	6.91042	28.80972	**83**	9.39780
18.43909	**34**	6.97953	28.98275	**84**	9.43539
18.70829	**35**	7.04730	29.15476	**85**	9.47268
18.97367	**36**	7.11379	29.32576	**86**	9.50969
19.23538	**37**	7.17905	29.49576	**87**	9.54640
19.49359	**38**	7.24316	29.66479	**88**	9.58284
19.74842	**39**	7.30614	29.83287	**89**	9.61900
20.00000	**40**	7.36806	30.00000	**90**	9.65489
20.24846	**41**	7.42896	30.16621	**91**	9.69052
20.49390	**42**	7.48887	30.33150	**92**	9.72589
20.73644	**43**	7.54784	30.49590	**93**	9.76100
20.97618	**44**	7.60591	30.65932	**94**	9.79586
21.21320	**45**	7.66309	30.82207	**95**	9.83048
21.44761	**46**	7.71944	30.98387	**96**	9.86485
21.67948	**47**	7.77498	31.14482	**97**	9.89898
21.90890	**48**	7.82974	31.30495	**98**	9.93288
22.13594	**49**	7.88374	31.46427	**99**	9.96656
22.36068	**50**	7.93701	31.62278	**100**	10.00000

TABLE CIX (Continued)
COMPUTATIONAL VALUES

$\sqrt{10n}$	n	$\sqrt[3]{100n}$	$\sqrt{10n}$	n	$\sqrt[3]{100n}$
10.0000	**10**	10.0000	71.4143	**510**	37.0843
14.1421	**20**	12.5992	72.1110	**520**	37.3251
17.3205	**30**	14.4225	72.8011	**530**	37.5629
20.0000	**40**	15.8740	73.4847	**540**	37.7976
22.3607	**50**	17.0998	74.1620	**550**	38.0295
24.4949	**60**	18.1712	74.8331	**560**	38.2586
26.4575	**70**	19.1293	75.4983	**570**	38.4850
28.2843	**80**	20.0000	76.1577	**580**	38.7088
30.0000	**90**	20.8008	76.8115	**590**	38.9300
31.6228	**100**	21.5444	77.4597	**600**	39.1487
33.1662	**110**	22.2398	78.1025	**610**	39.3650
34.6410	**120**	22.8943	78.7401	**620**	39.5789
36.0555	**130**	23.5134	79.3725	**630**	39.7906
37.4166	**140**	24.1014	80.0000	**640**	40.0000
38.7298	**150**	24.6621	80.6226	**650**	40.2073
40.0000	**160**	25.1984	81.2404	**660**	40.4124
41.2311	**170**	25.7128	81.8535	**670**	40.6155
42.4264	**180**	26.2074	82.4621	**680**	40.8166
43.5890	**190**	26.6840	83.0662	**690**	41.0157
44.7214	**200**	27.1442	83.6660	**700**	41.2129
45.8258	**210**	27.5892	84.2615	**710**	41.4082
46.9042	**220**	28.0204	84.8528	**720**	41.6017
47.9583	**230**	28.4387	85.4400	**730**	41.7934
48.9898	**240**	28.8450	86.0233	**740**	41.9834
50.0000	**250**	29.2402	86.6025	**750**	42.1716
50.9902	**260**	29.6250	87.1780	**760**	42.3582
51.9615	**270**	30.0000	87.7496	**770**	42.5432
52.9150	**280**	30.3659	88.3176	**780**	42.7266
53.8516	**290**	30.7232	88.8819	**790**	42.9084
54.7723	**300**	31.0723	89.4427	**800**	43.0887
55.6776	**310**	31.4138	90.0000	**810**	43.2675
56.5685	**320**	31.7480	90.5539	**820**	43.4448
57.4456	**330**	32.0753	91.1043	**830**	43.6207
58.3095	**340**	32.3961	91.6515	**840**	43.7952
59.1608	**350**	32.7107	92.1954	**850**	43.9683
60.0000	**360**	33.0193	92.7362	**860**	44.1401
60.8276	**370**	33.3222	93.2738	**870**	44.3105
61.6441	**380**	33.6198	93.8083	**880**	44.4796
62.4500	**390**	33.9121	94.3398	**890**	44.6475
63.2456	**400**	34.1995	94.8683	**900**	44.8141
64.0312	**410**	34.8422	95.3939	**910**	44.9794
64.8074	**420**	34.7603	95.9166	**920**	45.1436
65.5744	**430**	35.0340	96.4365	**930**	45.3066
66.3325	**440**	35.3035	96.9536	**940**	45.4684
67.0820	**450**	35.5689	97.4679	**950**	45.6290
67.8233	**460**	35.8305	97.9796	**960**	45.7886
68.5565	**470**	36.0883	98.4886	**970**	45.9470
69.2820	**480**	36.3424	98.9949	**980**	46.1044
70.0000	**490**	36.5931	99.4987	**990**	46.2607
70.7107	**500**	36.8403	100.0000	**1000**	46.4159

TABLE CX
EXPONENTIAL FUNCTIONS

e^x	x	e^{-x}	e^x	x	e^{-x}
1.0000	**0.0**	1.00000	148.41	**5.0**	0.00674
1.1052	**0.1**	0.90484	164.02	**5.1**	.00610
1.2214	**0.2**	.81873	181.27	**5.2**	.00552
1.3499	**0.3**	.74082	200.34	**5.3**	.00499
1.4918	**0.4**	.67032	221.41	**5.4**	.00452
1.6487	**0.5**	.60653	244.69	**5.5**	.00409
1.8221	**0.6**	.54881	270.43	**5-6**	.00370
2.0138	**0.7**	.49659	298.87	**5.7**	.00335
2.2255	**0.8**	.44933	330.30	**5.8**	.00303
2.4596	**0.9**	.40657	365.04	**5.9**	.00274
2.7183	**1.0**	.36788	403.43	**6.0**	.00248
3.0042	**1.1**	.33287	445.86	**6.1**	.00224
3.3201	**1.2**	.30119	492.75	**6.2**	.00203
3.6693	**1.3**	.27253	544.57	**6.3**	.00184
4.0552	**1.4**	.24660	601.85	**6.4**	.00166
4.4817	**1.5**	.22314	665.14	**6.5**	.00150
4.9530	**1.6**	.20190	735.10	**6.6**	.00136
5.4739	**1.7**	.18268	812.41	**6.7**	.00123
6.0496	**1.8**	.16530	897.85	**6.8**	.00111
6.6859	**1.9**	.14960	992.27	**6.9**	.00101
7.3891	**2.0**	.13534	1096.6	**7.0**	.00091
8.1662	**2.1**	.12246	1212.0	**7.1**	.00083
9.0250	**2.2**	.11080	1339.4	**7.2**	.00075
9.9742	**2.3**	.10026	1480.3	**7.3**	.00068
11.023	**2.4**	.09072	1636.0	**7.4**	.00061
12.182	**2.5**	.08209	1808.0	**7.5**	.00055
13.464	**2.6**	.07427	1998.2	**7.6**	.00050
14.880	**2.7**	.06721	2208.3	**7.7**	.00045
16.445	**2.8**	.06081	2440.6	**7.8**	.00041
18.174	**2.9**	.05502	2697.3	**7.9**	.00037
20.086	**3.0**	.04979	2981.0	**8.0**	.00034
22.198	**3.1**	.04505	3294.5	**8.1**	.00030
24.533	**3.2**	.04076	3641.0	**8.2**	.00028
27.113	**3.3**	.03688	4023.9	**8.3**	.00025
29.964	**3.4**	.03337	4447.1	**8.4**	.00023
33.115	**3.5**	.03020	4914.8	**8.5**	20010
36.598	**3.6**	.02732	5431.7	**8.6**	.00018
40.447	**3.7**	.02472	6002.9	**8.7**	.00017
44.701	**3.8**	.02237	6634.2	**8.8**	.00015
49.402	**3.9**	.02024	7332.0	**8.9**	.00014
54.598	**4.0**	.01832	8103.1	**9.0**	.00012
60.340	**4.1**	.01657	8955.3	**9.1**	.00011
66.686	**4.2**	.01500	9897.1	**9.2**	.00010
73.700	**4.3**	.01360	10938	**9.3**	.00009
81.451	**4.4**	.01228	12088	**9.4**	.00008
90.017	**4.5**	.01111	13360	**9.5**	.00008
99.632	**4.6**	.01005	14765	**9 6**	.00007
109.95	**4.7**	.00910	16318	**9.7**	.00006
121.51	**4.8**	.00823	18034	**9.8**	.00006
134.29	**4.9**	.00745	19930	**9.9**	.00005
			22026	**10.0**	.00005

TABLE CXI
STATISTICAL FORMULAS

Ia $M_A = \frac{\Sigma v}{n}$

Ib ${}_wM_A = \frac{\Sigma(wv)}{\Sigma w}$

Ic ${}_aM_A = \frac{\Sigma(fm)}{n}$

Id ${}_aM_A = \frac{\Sigma(fm)}{\Sigma f}$

IIa $M_d = \frac{n+1}{2},\ M_d = \frac{n}{2}$

IIb ${}_iM_d = L_1 + \frac{\frac{n}{2} - \Sigma_1}{f} \cdot i$

IIc ${}_iM_d = L_2 - \frac{\frac{n}{2} - \Sigma_2}{f} \cdot i$

III ${}_iM_o = L_m + \frac{f_h}{f_l + f_h} \cdot i$

IVa $M_g = \sqrt[n]{a_1 \cdot a_2 \cdot a_3 \cdot \ldots \cdot a_n}$

IVb $\log M_G = \frac{\log a_1 + \log a_2 + \log a_3 + \ldots + \log a_n}{n}$

IVc $\log {}_fM_G = \frac{f_1 \log m_1 + f_2 \log m_2 + f_3 \log m_3 + \ldots + f_n \log m_n}{\Sigma f}$

IVd $\log {}_fM_G = \frac{\Sigma(f \log m)}{\Sigma f}$

V $\frac{1}{M_H} = \frac{\Sigma\left(\frac{1}{r_1} + \frac{1}{r_2} + \frac{1}{r_3} + \ldots + \frac{1}{r_n}\right)}{n}$

VIa $M_Q = \sqrt{\frac{\Sigma(d)^2}{n}}$

VIb ${}_aM_Q = \sqrt{\frac{\Sigma f(d_a)^2}{n} - \left(\frac{\Sigma fd_a}{n}\right)^2}$

VIIa $R = a_{min} - a_{max}$

VIIb $R = a_{max} - a_{min}$

VIII $Q_2 = \frac{n}{2}$

IX $Q_1 = \frac{n}{4}$

X $Q_3 = \frac{3n}{4}$

XI $\mathrm{IQR} = Q_3 - Q_1$

XII $\mathrm{QDv} = \frac{Q_3 - Q_1}{2}$

XIII ${}_cD_{sp} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$

XIV ${}_cV = \frac{\sigma}{M_A} \cdot 100$

XV $D_1 = \frac{n}{10}$

XVI $D_9 = \frac{9n}{10}$

XVII $D_5 = \frac{n}{2}$

XVIII $\mathrm{PCR} = D_9 - D_1$

XIXa $\mathrm{Dv}_{M_A} = \frac{\Sigma[x]}{n}$

XIXb $\mathrm{Dv}_{M_d} = \frac{\Sigma[d]}{n}$

XIXc ${}_f\mathrm{Dv}_{M_A} = \frac{\Sigma(f[m_x])}{n}$

XIXd ${}_f\mathrm{Dv}_{M_A} = \frac{\Sigma(f[m_d])}{n}$

XX $\sigma = \sqrt{\frac{\Sigma(d_m)^2}{n}}$

XXI $c = \frac{\Sigma fd_m}{\Sigma f}$

XXII $\sigma = i\sqrt{\frac{\Sigma f(d_m)^2}{\Sigma f} - c^2}$

XXIII $\Sigma f(d_m+1)^2 = \Sigma f(d_m)^2 + 2\Sigma fd_m + \Sigma f$

XXIV $M_o = M_A - 3(M_A - M_d)$

XXV $\mathrm{Sk} = \frac{M_A - M_d}{\sigma}$

XXVI $y_c = a + bx$

XXVIIa $\Sigma y = na + b\Sigma x$

XXVIIb $\Sigma y = na$

XXVIIc $\Sigma \log y = n \log a + \log b\, \Sigma x$

XXVIId $\log a = \frac{\Sigma \log y}{n}$

XXVIIIa $\Sigma(xy) = a\Sigma x + b\Sigma(x^2)$

XXVIIIb $\Sigma(xy) = b\Sigma(x^2)$

XXVIIIc $\Sigma(x \log y) = \log a\, \Sigma x + \log b\, \Sigma(x^2)$

XXVIIId $\log b = \frac{\Sigma(x \log y)}{\Sigma(x^2)}$

XXIXa $\log y_c = \log a + (\log b)x$

XXIXb $y_c = a(b)^x$

XXX $y_c = a + bx + c^2$

XXXI $y_c = a + bx + cx^2 + dx^3$

XXXIIa $y_c = ab^{c^x}$

XXXIIb $\log y_c = \log a + (\log b)c^x$

XXXIII $\frac{1}{y_c} = a + bc^x$

XXXIVa $O = T \times S \times C \times I$

XXXIVb $O = T \times C \times I$

TABLE CXI (Continued)

STATISTICAL FORMULAS

XXXIVc $\frac{O}{T}=C\times I$

XXXV $\sigma=\sqrt{\frac{\Sigma(d^2)}{n}-\left(\frac{\Sigma d}{n}\right)^2}$

XXXVI $\sigma_{M_A}=\frac{\sigma_S}{\sqrt{n-1}}$

XXXVII $t=\frac{k_S-k_U}{\frac{\sigma}{\sqrt{n-m+1}}}$

XXXVIII $\sigma_{D_{1-2}}=\sqrt{\sigma^2_{M_{A_1}}+\sigma^2_{M_{A_2}}}$

XXXIX $T=\frac{M_{A_1}-M_{A_2}}{\sigma_{D_{1-2}}}$

XL $s=\sqrt{\frac{\Sigma(x_1^2)+\Sigma(x_2^2)}{(n_1-1)+(n_2-1)}}$

XLI $\sigma_{M_d}=1.2530\cdot\frac{\sigma}{n}$

XLII $\sigma_\sigma=\frac{\sigma}{\sqrt{2n}}$

XLIII $\sigma_V=\frac{V}{\sqrt{2n}}\sqrt{1+\frac{2(V^2)}{10^4}}$

XLIV $V=\frac{\sigma}{M_A}\cdot 100$

XLV $PE=0.6745\sigma$

XLVI $PE_{M_A}=0.6745\cdot\frac{\sigma}{\sqrt{n-1}}$

XLVII $PE_{M_d}=0.85535\cdot\frac{\sigma}{\sqrt{n}}$

XLVIII $PE\sigma=0.6745\cdot\frac{\sigma}{\sqrt{2n}}$

XLIX $PE_V=0.6745\cdot\frac{V}{\sqrt{2n}}\sqrt{1+\frac{2(V^2)}{10^4}}$

TABLE CXII

INDEX NUMBERS

I $\frac{\Sigma p_1}{\Sigma p_0}=\frac{p_1+p_1'+p_1''+\ldots+p_1^n}{p_0+p_0'+p_0''+\ldots+p_0^n}$

II $\frac{\Sigma\left(\frac{p_1}{p_0}\right)}{n}=\frac{p_1}{p_0}+\frac{p_1'}{p_0'}+\frac{p_1''}{p_0''}\ \ldots+\frac{p_1^n}{p_0^n}$

III $\frac{\Sigma p_1q_0}{\Sigma p_0q_0}=\frac{p_1q_0}{p_0q_0}+\frac{p_1'q_0'}{p_0'q_0'}+\frac{p_1''q_0''}{p_0''q_0''}+\ldots+\frac{p_1^nq_0^n}{p_0^nq_0^n}$

IV $\frac{\Sigma p_1q_0}{\Sigma p_0q_0}=\frac{p_1}{p_0}\cdot p_0q_0+\frac{p_1'}{p_0'}\cdot p_0'q_0'+\frac{p_1''}{p_0''}\cdot p_0''q_0''+\ldots+\frac{p_1^n}{p_0^n}\cdot p_0^nq_0^n$

V $\frac{\Sigma p_1q_1}{\Sigma p_0q_0}=\frac{p_1q_1}{p_0q_0}+\frac{p_1'q_1'}{p_0'q_0'}+\frac{p_1''q_1''}{p_0''q_0''}+\ldots+\frac{p_1^nq_1^n}{p_0^nq_0^n}$

VI Ideal Index Number $=\sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0}\times\frac{\Sigma p_1q_1}{\Sigma p_0q_0}}$

VII $\frac{\Sigma\left(\frac{q_1}{q_0}\right)}{n}$

VIII $\frac{\Sigma p_0q_1}{\Sigma p_0q_0}=\frac{p_0q_1}{p_0q_0}+\frac{p_0'q_1'}{p_0'q_0'}+\frac{p_0''q_1''}{p_0''q_0''}+\ldots+\frac{p_0^nq_1^n}{p_0^nq_0^n}$

IX $\frac{q_1}{q_0}\cdot p_0q_0+\frac{q_1'}{q_0'}\cdot p_0'q_0'+\frac{q_1''}{q_0''}\cdot p_0''q_0''+\ldots+\frac{q_1^n}{q_0^n}\cdot p_0^n\,q_0^n$

X $I_i=\frac{\Sigma p_1q_0}{\Sigma p_{i-1}q_0}\cdot I_{i-1}$ (i = current month)

TABLE CXIII

REGRESSION AND CORRELATION FORMULAS

I $y_c=a+bx$

IIa $a=\dfrac{\Sigma x^2\Sigma y-\Sigma x\,\Sigma xy}{n\Sigma x^2-(\Sigma x)^2}$

IIb $a=\dfrac{\Sigma y}{n}$

IIIa $b=\dfrac{n\Sigma xy-\Sigma x\Sigma y}{n\Sigma x^2-(\Sigma x)^2}$

IIIb $b=\dfrac{\Sigma xy}{\Sigma x^2}$

IVa $S_y=\sqrt{\dfrac{\Sigma(y-y_c)^2}{n}}$

IVb $S_y=\sqrt{\dfrac{\Sigma(y^2)-(a\Sigma y+b\Sigma xy)}{n}}$

V $\bar{S}_y=S_y^2\left(\dfrac{n}{n-m}\right)$

VI $r=\sqrt{1-\dfrac{(S_y)^2}{(\sigma_y)^2}}$

VII $r^2=1-\dfrac{(S_y)^2}{(\sigma_y)^2}$

VIII $k=\dfrac{(S_y)}{(\sigma_y)}$

IX $\rho_{yx}=\sqrt{1-\dfrac{(S_y)^2}{(\sigma_y)^2}}$

X $\bar{r}^2=1-(1-r^2)\left(\dfrac{n-1}{n-m}\right)$

XI $\bar{\rho}_{yx}^2=1-(1-\rho_{yx}^2)\left(\dfrac{n-1}{n-m}\right)$

XII $r=\dfrac{\rho}{\sigma_x\sigma_y}$

XIIIa $y=r\cdot\dfrac{\sigma_y}{\sigma_x}\cdot x$

XIIIb $x=r\cdot\dfrac{\sigma_x}{\sigma_y}\cdot y$

XIV $\rho_S=1-\dfrac{6\Sigma(|D|^2)}{n(n^2-1)}$

XV $r=2\sin\left(\dfrac{\pi}{6}\rho\right)$

XVI $R=1-\dfrac{6\Sigma G}{n^2-1}$

XVII $r=\dfrac{\Sigma d_x d_y}{\sqrt{\Sigma(d_x)^2\cdot\Sigma(d_y)^2}}$

XVIII $x_1=a+b_{12.3}x_2+b_{13.2}x_3$

XIX $x_1=a+b_{12.34}x_2+b_{13.24}x_3+b_{14.23}x_4$

XXa $S_{1.234}\sqrt{\dfrac{\Sigma(d^2)}{n}}$

XXb $S_{1.234}^2=\sigma_1^2-b_{12.34}p_{12}-b_{13.24}p_{13}-b_{14.23}p_{14}$

XXIa $R_{1.234}=\sqrt{1-\dfrac{S_{1.234}^2}{\sigma_1^2}}$

XXIb $R_{1.234}^2=\dfrac{b_{12.34}p_{12}+b_{13.24}p_{13}+b_{14.23}p_{14}}{\sigma_1^2}$

XXII $x_1=a+fx_2+fx_3+fx_4+\ldots$

XXIIIa $r_{12.3}=\dfrac{r_{12}-r_{13}r_{23}}{(\sqrt{1-r_{13}^2})(\sqrt{1-r_{23}^2})}$

XXIIIb $r_{13.24}=\sqrt{b_{13.24}\cdot b_{31.24}}$

XXIIIc $r_{14.23}^2=1-\dfrac{1-R_{1.234}^2}{1-R_{1.23}^2}$

or $r_{14.23}=\sqrt{\dfrac{1-S_{1.234}^2}{S_{1.23}^2}}$

XXIV ${}_{12}r_{34}=\dfrac{b_{12.34}^2\cdot\sigma_2^2}{b_{12.34}^2\sigma_2^2+\sigma_1^2(1-R_{1.234}^2)}$

XXVa $\beta_{12.3}=b_{12.3}\cdot\dfrac{\sigma_2}{\sigma_1}$

XXVb $\beta_{13.2}=b_{13.2}\cdot\dfrac{\sigma_3}{\sigma_1}$

TABLE CXIV

NAPERIAN (NATURAL) LOGARITHMS

The first ten rows provide logarithms of numbers from 0.0 to 9.9, at intervals of one-tenth (0.1).
The remainder of the table presents logarithms from 10 to 1009 at intervals of one (1).
The natural logarithm of 1.1 is found by looking in the row beginning 1. under the column headed 1: 0.09531.
The natural logarithm of 11 is found by looking in the row beginning 1 under the column headed 1: 2.39790.

	0	1	2	3	4	5	6	7	8	9
0.		7.697–10	8.391–10	8.796–10	9.084–10	9.307–10	9.489–10	9.643–10	9.777–10	9.895–10
1.	0.00000	0.09531	0.18232	0.26236	0.33647	0.40547	0.47000	0.53063	0.58779	0.64185
2.	0.69315	0.74194	0.78846	0.83291	0.87547	0.91629	0.95551	0.99325	1.02962	1.06471
3.	1.09861	1.13140	1.16315	1.19392	1.22378	1.25276	1.28093	1.30833	1.33500	1.36098
4.	1.38629	1.41099	1.43508	1.45862	1.48160	1.50408	1.52606	1.54756	1.56862	1.58924
5.	1.60944	1.62924	1.64866	1.66771	1.68640	1.70475	1.72277	1.74047	1.75786	1.77495
6.	1.79176	1.80829	1.82455	1.84055	1.85630	1.87180	1.88707	1.90211	1.91692	1.93152
7.	1.94591	1.96009	1.97408	1.98787	2.00148	2.01490	2.02815	2.04122	2.05412	2.06686
8.	2.07944	2.09186	2.10413	2.11626	2.12823	2.14007	2.15176	2.16332	2.17475	2.18605
9.	2.19722	2.20827	2.21920	2.23001	2.24071	2.25129	2.26176	2.27213	2.28238	2.29253
1	2.30259	2.39790	2.48491	2.56495	2.63906	2.70805	2.77259	2.83321	2.89037	2.94444
2	2.99573	3.04452	3.09104	3.13549	3.17805	3.21888	3.25810	3.29584	3.33220	3.36730
3	3.40120	3.43399	3.46574	3.49651	3.52636	3.55535	3.58352	3.61092	3.63759	3.66356
4	3.68888	3.71357	3.73767	3.76120	3.78419	3.80666	3.82864	3.85015	3.87120	3.89182
5	3.91202	3.93183	3.95124	3.97029	3.98898	4.00733	4.02535	4.04305	4.06044	4.07754
6	4.09434	4.11087	4.12713	4.14313	4.15888	4.17439	4.18965	4.20469	4.21951	4.23411
7	4.24850	4.26268	4.27667	4.29046	4.30407	4.31749	4.33073	4.34381	4.35671	4.36945
8	4.38203	4.39445	4.40672	4.41884	4.43082	4.44265	4.45435	4.46591	4.47734	4.48864
9	4.49981	4.51086	4.52179	4.53260	4.54329	4.55388	4.56435	4.57471	4.58497	4.59512
10	4.60517	4.61512	4.62497	4.63473	4.64439	4.65396	4.66344	4.67283	4.68213	4.69135
11	4.70048	4.70953	4.71850	4.72739	4.73620	4.74493	4.75359	4.76217	4.77068	4.77912
12	4.78749	4.79579	4.80402	4.81218	4.82028	4.82831	4.83628	4.84419	4.85203	4.85981
13	4.86753	4.87520	4.88280	4.89035	4.89784	4.90527	4.91265	4.91998	4.92725	4.93447
14	4.94164	4.94876	4.95583	4.96284	4.96981	4.97673	4.98361	4.99043	4.99721	5.00395
15	5.01064	5.01728	5.02388	5.03044	5.03695	5.04343	5.04986	5.05625	5.06260	5.06890
16	5.07517	5.08140	5.08760	5.09375	5.09987	5.10595	5.11199	5.11799	5.12396	5.12990
17	5.13580	5.14166	5.14749	5.15329	5.15906	5.16479	5.17048	5.17615	5.18178	5.18739
18	5.19296	5.19850	5.20401	5.20949	5.21494	5.22036	5.22575	5.28111	5.23644	5.24175
19	5.24702	5.25227	5.25750	5.26269	5.26786	5.27300	5.27811	5.28320	5.28827	5.29330
20	5.29832	5.30330	5.30827	5.31321	5.31812	5.32301	5.32788	5.33272	5.33754	5.34233
21	5.34711	5.35186	5.35659	5.36129	5.36598	5.37064	5.37528	5.37990	5.38450	5.38907
22	5.39363	5.39816	5.40268	5.40717	5.41165	5.41610	5.42053	5.42495	5.42935	5.43372
23	5.43808	5.44242	5.44674	5.45104	5.45532	5.45959	5.46383	5.46806	5.47227	5.47646
24	5.48064	5.48480	5.48894	5.49306	5.49717	5.50126	5.50533	5.50939	5.51343	5.51745
25	5.52146	5.52545	5.52943	5.53339	5.53733	5.54126	5.54518	5.54908	5.55296	5.55683
26	5.56068	5.56452	5.56834	5.57215	5.57595	5.57973	5.58350	5.58725	5.59099	5.59471
27	5.59842	5.60212	5.60580	5.60947	5.61313	5.61677	5.62040	5.62402	5.62762	5.63121
28	5.63479	5.63835	5.64191	5.64545	5.64897	5.65249	5.65599	5.65948	5.66296	5.66643
29	5.66988	5.67332	5.67675	5.68017	5.68358	5.68698	5.69036	5.69373	5.69709	5.70044
30	5.70378	5.70711	5.71043	5.71373	5.71703	5.72031	5.72359	5.72685	5.73010	5.73334
31	5.73657	5.73979	5.74300	5.74620	5.74939	5.75257	5.75574	5.75890	5.76205	5.76519
32	5.76832	5.77144	5.77455	5.77765	5.78074	5.78383	5.78690	5.78996	5.79301	5.79606
33	5.79909	5.80212	5.80513	5.80814	5.81114	5.81413	5.81711	5.82008	5.82305	5.82600
34	5.82895	5.83188	5.83481	5.83773	5.84064	5.84354	5.84644	5.84932	5.85220	5.85507
35	5.85793	5.86079	5.86363	5.86647	5.86930	5.87212	5.87493	5.87774	5.88053	5.88332
36	5.88610	5.88888	5.89164	5.89440	5.89715	5.89990	5.90263	5.90536	5.90808	5.91080
37	5.91350	5.91620	5.91889	5.92158	5.92426	5.92693	5.92959	5.93225	5.93489	5.93754
38	5.94017	5.94280	5.94542	5.94803	5.95064	5.95324	5.95584	5.95842	5.96101	5.96358
39	5.96615	5.96871	5.97126	5.97381	5.97635	5.97889	5.98141	5.98394	5.98645	5.98896
40	5.99146	5.99396	5.99645	5.99894	6.00141	6.00389	6.00635	6.00881	6.01127	6.01372
41	6.01616	6.01859	6.02102	6.02345	6.02587	6.02828	6.03069	6.03309	6.03548	6.03787
42	6.04025	6.04263	6.04501	6.04737	6.04973	6.05209	6.05444	6.05678	6.05912	6.06146
43	6.06379	6.06611	6.06843	6.07074	6.07304	6.07535	6.07764	6.07993	6.08222	6.08450
44	6.08677	6.08904	6.09131	6.09357	6.09582	6 09807	6.10032	6.10256	6.10479	6.10702
45	6.10925	6.11147	6.11368	6.11589	6.11810	6.12030	6.12249	6.12468	6.12687	6.12905

TABLE CXIV (Continued)
NAPERIAN (NATURAL) LOGARITHMS

To find the natural logarithm of a number ten times as large as the given number, add to the given logarithm the logarithm of 10.

To find the natural logarithm of a number one-tenth as large as the given number, subtract from the given logarithm the logarithm of 10.

	0	1	2	3	4	5	6	7	8	9
46	6.13123	6.13340	6.13556	6.13773	6.13988	6.14204	6.14419	6.14633	6.14847	6.15060
47	6.15273	6.15486	6.15698	6.15910	6.16121	6.16331	6.16542	6.16752	6.16961	6.17170
48	6.17379	6.17587	6.17794	6.18002	6.18208	6.18415	6.18621	6.18826	6.19032	6.19236
49	6.19441	6.19644	6.19848	6.20051	6.20254	6.20456	6.20658	6.20859	6.21060	6 21261
50	6.21461	6.21661	6.21860	6.22059	6.22258	6.22456	6.22654	6.22851	6.23048	6.23245
51	6.23441	6.23637	6.23832	6.24028	6.24222	6.24417	6.24611	6.24804	6.24998	6.25190
52	6.25383	6.25575	6.25767	6.25958	6.26149	6.26340	6.26530	6.26720	6.26910	6.27099
53	6.27288	6.27476	6.27664	6.27852	6.28040	6.28227	6.28413	6.28600	6.28786	6.28972
54	6.29157	6.29342	6.29527	6.29711	6.29895	6.30079	6.30262	6.30445	6.30628	6.30810
55	6.30992	6.31173	6.31355	6.31536	6.31716	6.31897	6.32077	6.32257	6.32436	6.32615
56	6.32794	6.32972	6.33150	6.33328	6.33505	6.33683	6.33859	6.34036	6.34212	6.34388
57	6.34564	6.34739	6.34914	6.35089	6.35263	6.35437	6.35611	6.35784	6.35957	6.36130
58	6.36303	6.36475	6.36647	6.36819	6.36990	6.37161	6.37332	6.37502	6.37673	6.37843
59	6.38012	6.38182	6.38351	6.38519	6.38688	6.38856	6.39024	6.39192	6.39359	6.39526
60	6.39693	6.39859	6.40026	6.40192	6.40357	6.40523	6.40688	6.40853	6.41017	6.41182
61	6.41346	6.41510	6.41673	6.41836	6.41999	6.42162	6.42325	6.42487	6.42649	6.42811
62	6.42972	6.43133	6.43294	6.43455	6.43615	6.43775	6.43935	6.44095	6.44254	6.44413
63	6.44572	6.44731	6.44889	6.45047	6.45205	6.45362	6.45520	6.45677	6.45834	6.45990
64	6.46147	6.46303	6.46459	6.46614	6.46770	6.46925	6.47080	6.47235	6.47389	6.47543
65	6.47697	6.47851	6.48004	6.48158	6.48311	6.48464	6.48616	6.48768	6.48920	6.49072
66	6.49224	6.49375	6.49527	6.49677	6.49828	6.49979	6.50129	6.50279	6.50429	6.50578
67	6.50728	6.50877	6.51026	6.51175	6.51323	6.51471	6.51619	6.51767	6.51915	6.52062
68	6.52209	6.52356	6.52503	6.52649	6.52796	6.52942	6.53088	6.53233	6.53379	6.53524
69	6.53669	6.53814	6.53959	6.54103	6.54247	6.54391	6.54535	6.54679	6.54822	6.54965
70	6.55108	6.55251	6.55393	6.55536	6.55678	6.55820	6.55962	6.56103	6.56244	6.56386
71	6.56526	6.56667	6.56808	6.56948	6.57088	6.57228	6.57368	6.57508	6.57647	6.57786
72	6.57925	6.58064	6.58203	6.58341	6.58479	6.58617	6.58755	6.58893	6.59030	6.59167
73	6.59304	6.59441	6.59578	6.59715	6.59851	6.59987	6.60123	6.60259	6.60394	6.60530
74	6.60665	6.60800	6.60935	6.61070	6.61204	6.61338	6.61473	6.61607	6.61740	6.61874
75	6.62007	6.62141	6.62274	6.62407	6.62539	6.62672	6.62804	6.62936	6.63068	6.63200
76	6.63332	6.63463	6.63595	6.63726	6.63857	6.63988	6.64118	6.64249	6.64379	6.64509
77	6.64639	6.64769	6.64898	6.65028	6.65157	6.65286	6.65415	6.65544	6.65673	6.65801
78	6.65929	6.66058	6.66185	6.66313	6.66441	6.66568	6.66696	6.66823	6 66950	6.67077
79	6.67203	6.67330	6.67456	6.67582	6.67708	6.67834	6.67960	6.68085	6.68211	6.68336
80	6.68461	6.68586	6.68711	6.68835	6.68960	6.69084	6.69208	6.69332	6.69456	6.69580
81	6.69703	6.69827	6.69950	6.70073	6.70196	6.70319	6.70441	6.70564	6.70686	6.70808
82	6.70930	6.71052	6.71174	6.71296	6.71417	6.71538	6.71659	6.71780	6.71901	6.72022
83	6.72143	6.72263	6.72383	6.72503	6.72623	6.72743	6.72863	6.72982	6.73102	6.73221
84	6.73340	6.73459	6.73578	6.73697	6.73815	6.73934	6.74052	6.74170	6.74288	6.74406
85	6.74524	6.74641	6.74759	6.74876	6.74993	6.75110	6.75227	6.75344	6.75460	6.75577
86	6.75693	6.75809	6.75926	6.76401	6.76157	6.76273	6.76388	6.76504	6.76619	6.76734
87	6.76849	6.76964	6.77079	6.77194	6.77308	6.77422	6.77537	6 77651	6.77765	6 77878
88	6.77922	6.78106	6.78219	6.78333	6.78446	6.78559	6.78672	6.78784	6.78897	6.79010
89	6.79122	6.79234	6.79347	6.79459	6.79571	6.79682	6.79794	6.79906	6.80017	6.80128
90	6.80239	6.80351	6.80461	6.80572	6.80683	6.80793	6.80904	6.81014	6.81124	6.81235
91	6.81344	6.81454	6.81564	6.81674	6.81783	6.81892	6.82002	6.82111	6.82220	6.82329
92	6.82437	6.82546	6.82655	6.82763	6.82871	6.82979	6.83087	6.83195	6.83303	6.83411
93	6.83518	6.83626	6.83733	6.83841	6.83948	6.84055	6.84162	6.84268	6.84375	6.84482
94	6.84588	6.84694	6.84801	6.84907	6.85013	6.85118	6.85224	6.85330	6.85435	6.85541
95	6.85646	6.85751	6.85857	6.85961	6.86066	6.86171	6.86276	6.86380	6.86485	6.86589
96	6.86693	6.86797	6.86901	6.87005	6.87109	6.87213	6.87316	6.87420	6.87523	6.87626
97	6.87730	6.87833	6.87936	6.88038	6.88141	6.88244	6.88346	6.88449	6.88551	6.88653
98	6.88755	6.88857	6.88959	6.89061	6.89163	6.89264	6.89366	6.89467	6.89568	6.89669
99	6.89770	6.89871	6.89972	6.90073	6.90174	6.90274	6.90375	6.90475	6.90575	6.90675
100	6.90076	6.90875	6.90975	6.91075	6.91175	6.91274	6.91374	6.91473	6.91572	6.91672

Solutions to Problems and Exercises

Answers achieved by readers may differ by minor amounts in some instances if different tables, different methods of calculation, or retention of a different number of decimal places happen to be employed. Such discrepancies, if they do not occur in the significant figures (cf. p. 68), should not be a cause of concern.

FUNDAMENTALS OF STATISTICS

The nature of the answers in this section makes it seem desirable to reproduce parts of the worksheets from which the values for the formulas are obtained. To save space, these data are arranged in parallel columns in one tabular arrangement for each table. The bold numerals preceding the column headings indicate the problems for which the columns provide materials used in the solutions.

WORK-SHEET FOR TABLE α

	1	**4** Cumulated Frequencies		**5**	**9**
	f	<	>	fc	log a
398	5	5	120	1990	2.59988
399	10	15	115	3990	2.60097
400	16	31	105	6400	2.60206
401	23	54	89	9223	2.60314
402	20	74	66	8040	2.60423
403	17	91	46	6851	2.60531
404	13	104	29	5252	2.60638
405	9	113	16	3645	2.60746
406	4	117	7	1624	2.60853
407	2	119	3	814	2.60959
408	1	120	1	408	2.61066
$n=11$	$\Sigma f=120$			$\Sigma fc=48237$	Σ log $a=28.65821$
					$a=$ 2.60522

WORK-SHEET FOR TABLE β

i	m	**2** f	**4** Cumulated Frequencies <	>	**5** fm	**9** log m	f log m
4.375-4.625	4.50	8	8	120	36.00	0.65321	5.22568
4.625-4.875	4.75	10	18	112	47.50	0.67669	6.76690
4.875-5.125	5.00	12	30	102	60.00	0.69897	8.38764
5.125-5.375	5.25	59	89	90	309.75	0.72016	42.48944
5.375-5.625	5.50	15	104	31	82.50	0.74036	11.10540
5.625-5.875	5.75	8	112	16	46.00	0.75967	6.07736
5.875-6.125	6.00	8	120	8	48.00	0.77085	6.16680
		$\Sigma f=120$					Σf log $m=86.21922$
							0.71933

ANSWERS

For convenience of reference, problem numbers are keyed with the Greek letters indicating the tables on which the work is based.

1 α See col. 2, α work-sheet
2 β See col. 2, β work-sheet
3 Graphic solution
4 α, β See cols. 3 and 4, α and β work-sheets
5 α 401.14 β 5.24783
6 α 401.275 β 5.254
7 α 401
8 β (a) 5.125-5.375 (b) 5.254
9 α 402.9 β 5.24
10 γ 11058.11 **11** δ 9.17405
12 $M_d=$ 869 $M_A=867$ $M_G=863.34$

13 α 398-408 β 4.45-6.05
14 α 2 β 0.2425
15 α 6.1 β 1.075
16 α 1.811 β 0.2316
17 α 2.1514 β 0.288425
18 α 401.545 β 5.26634
19 α −0.0626 β 0.02139
20 ε

1931	416,648	1936	784,587
1932	235,187	1937	893,085
1933	346,545	1938	488,100
1934	575,192	1939	710,496
1935	694,690	1940	777,026

	Year	Jan.	Feb.	March	April	May	June	July	August	Sept.	Oct.	Nov.	Dec.
21 ε	1931	8.0	9.5	10.8	12.0	11.0	9.7	8.2	7.6	7.5	5.2	4.7	5.7
	1932	8.7	9.9	8.3	11.2	11.3	9.7	6.1	6.1	8.2	5.8	5.1	9.0
	1933	5.5	4.4	5.1	7.7	9.7	12.2	11.0	12.0	9.9	8.6	5.3	8.6
	1934	7.5	7.6	10.3	11.2	9.9	7.9	7.3	8.9	7.8	8.3	6.0	7.4
	1935	8.9	8.4	9.6	9.5	8.0	8.9	8.3	8.1	4.5	8.5	8.4	8.9
	1936	8.6	8.1	9.9	11.0	9.6	9.9	8.8	7.9	5.7	4.4	6.9	9.5
	1937	7.9	7.5	10.1	10.8	10.2	9.6	8.8	9.3	5.9	3.5	7.2	9.2
	1938	11.0	9.7	9.8	8.8	7.6	7.8	7.1	6.5	3.8	4.5	10.7	12.8
	1939	8.5	8.4	10.2	9.0	8.4	8.9	8.3	5.4	3.8	8.7	9.4	11.0
	1940	9.0	8.6	9.1	9.1	8.4	7.5	8.1	3.7	5.7	9.3	10.3	11.2

	Year	Jan.	Feb.	March	April	May	June	July	August	Sept.	Oct.	Nov.	Dec.
22 *	1931	84.0	109.2	140.4	174.0	165.6	128.4	111.6	94.8	66.0	36.0	30.0	54.0
	1932	104.4	99.6	105.6	128.4	166.8	169.2	99.6	80.4	67.2	37.2	50.4	91.2
	1933	84.0	78.4	74.4	114.0	138.0	158.4	146.4	146.4	120.0	80.4	32.4	38.4
	1934	62.4	102.0	153.6	158.4	151.2	138.0	122.4	100.8	68.4	46.8	27.6	61.2
	1935	84.0	100.8	133.2	142.8	112.8	108.0	100.8	67.2	20.4	79.2	124.8	126.0
	1936	97.2	73.2	111.6	135.6	126.0	122.4	121.2	67.4	30.0	62.4	111.6	139.2
	1937	94.8	91.2	123.6	134.4	122.4	126.0	110.4	96.0	36.0	91.2	90.0	75.6
	1938	93.6	84.0	104.4	105.6	92.4	81.6	63.6	34.8	39.6	112.8	192.0	195.6
	1939	117.6	102.0	126.0	114.0	99.6	103.2	63.6	25.2	67.2	105.6	118.8	156.0
	1940	117.6	109.2	115.2	117.6	106.8	92.4	55.2	15.6	73.2	136.8	132.0	128.4

	Year	Jan.	Feb.	March	April	May	June	July	August	Sept.	Oct.	Nov.	Dec.
23 ε	1931	96.0	114.0	129.6	144.0	132.0	116.4	98.4	91.2	90.0	62.4	56.4	68.4
	1932	104.4	118.8	99.6	134.4	135.6	116.4	73.2	73.2	98.4	69.6	61.2	108.0
	1933	66.0	52.8	61.2	92.4	116.4	146.4	132.0	144.0	118.8	103.2	63.6	103.2
	1934	90.0	91.2	123.6	134.4	118.8	94.8	87.6	106.8	93.6	99.6	72.0	88.8
	1935	106.8	100.8	115.2	114.0	96.0	106.8	99.6	97.2	54.0	102.0	100.8	106.8
	1936	103.2	97.2	118.8	132.0	115.2	118.8	105.6	94.8	68.4	52.8	82.8	114.0
	1937	94.8	90.0	121.2	129.6	122.4	115.2	105.6	111.6	70.8	42.0	86.4	110.4
	1938	132.0	116.4	117.6	105.6	91.2	93.6	85.2	78.0	45.6	54.0	128.4	153.6
	1939	102.0	100.8	122.4	108.0	100.8	106.8	99.6	64.8	45.6	104.4	112.8	132.0
	1940	108.0	103.2	109.2	109.2	100.8	90.0	97.2	44.4	68.4	111.6	123.6	134.4

	Year	Jan.	Feb.	March	April	May	June	July	August	Sept.	Oct.	Nov.	Dec.
24 *	1931	82.4	118.6	137.8	176.4	162.5	130.2	109.5	93.0	66.9	35.3	30.4	53.0
	1932	102.7	104.7	103.9	130.5	164.1	172.0	97.9	79.1	68.3	36.6	51.2	90.4
	1933	82.4	87.2	73.0	115.5	135.4	160.6	143.6	143.6	121.6	78.9	32.9	37.7
	1934	61.2	110.8	150.7	160.6	148.3	139.9	120.1	98.9	69.4	45.9	28.0	60.1
	1935	82.4	109.5	130.7	144.8	110.7	109.5	98.9	65.9	20.7	77.7	126.5	123.6
	1936	95.6	77.0	109.8	137.9	123.9	124.4	119.2	66.3	30.5	61.4	113.5	136.9
	1937	93.0	99.1	121.3	136.2	120.1	127.8	108.3	94.2	36.5	89.5	91.3	74.2
	1938	91.8	91.2	102.4	107.1	90.7	82.7	62.4	34.1	40.2	110.7	194.6	191.9
	1939	115.4	110.8	123.6	115.6	97.7	104.6	62.4	24.8	68.1	103.6	120.5	153.1
	1940	115.7	114.8	113.3	119.6	105.0	93.9	54.2	15.3	74.4	134.6	134.2	126.3

	Year	Jan.	Feb.	March	April	May	June	July	August	Sept.	Oct.	Nov.	Dec.
25 ε	1931	94.2	123.8	127.0	145.9	129.4	118.0	96.4	89.4	91.3	61.2	57.2	67.0
	1932	102.7	124.9	96.0	136.7	133.4	118.3	72.0	72.0	110.1	68.5	62.2	106.3
	1933	64.7	57.3	60.1	93.7	114.2	148.4	129.5	141.3	120.5	101.3	64.5	101.3
	1934	88.3	99.1	121.3	136.3	116.5	96.1	85.8	104.8	94.9	97.7	73.0	87.1
	1935	104.8	109.5	113.0	115.6	94.2	108.3	97.7	95.4	54.7	101.1	102.2	104.8
	1936	101.5	102.2	116.8	134.2	113.3	120.7	103.9	93.2	69.5	51.9	84.2	112.1
	1937	92.0	97.8	118.9	131.4	120.1	116.8	103.6	109.5	71.8	41.2	87.6	109.3
	1938	129.5	126.4	115.4	107.1	89.5	94.9	83.6	76.5	46.2	53.0	130.2	150.7
	1939	100.1	109.5	120.1	109.5	98.9	108.3	97.7	63.6	46.2	102.4	114.4	129.5
	1940	106.2	108.5	107.4	111.0	99.1	91.5	95.6	43.7	69.6	109.8	125.7	132.2

26 ε 453,652.4-730658.8

27

(a)		(b)		(c)	
1931–1932	325,918	1931–1933	332,793	1931–1934	393,393
1932–1933	290,866	1932–1934	385,641	1932–1935	462,904
1933–1934	460,869	1933–1935	538,809	1933–1936	600,254
1934–1935	634,941	1934–1936	684,823	1934–1937	736,889
1935–1936	739,639	1935–1937	790,787	1935–1938	715,116
1936–1937	838,836	1936–1938	721,924	1936–1939	719,067
1937–1938	690,593	1937–1939	697,227	1937–1940	717,177
1938–1939	599,298	1938–1940	658,541		
1939–1940	743,761				

* Table 22.

	Year	Jan.	Feb.	March	April	May	June	July	August	Sept.	Oct.	Nov.	Dec.
28 *	1931	....	....	....	....	....	....	113.2	98.7	73.6	42.9	39.5	83.3
	1932	88.4	89.9	99.9	124.9	164.9	168.6	99.6	79.6	68.1	36.5	48.3	85.0
	1933	102.3	77.6	78.9	113.9	134.5	156.6	145.7	141.4	107.1	65.4	24.9	28.8
	1934	62.9	103.3	156.0	163.1	161.1	146.0	119.2	94.3	62.0	40.2	22.8	51.1
	1935	103.9	122.8	163.6	178.9	127.5	112.6	100.1	65.9	20.6	78.2	121.6	120.9
	1936	102.2	75.6	114.6	139.0	128.1	124.1	121.4	67.6	28.6	59.7	106.1	130.9
	1937	95.0	90.1	120.6	129.1	123.9	123.2	112.7	101.5	40.8	110.5	115.1	108.4
	1938	76.5	76.0	102.8	108.4	97.6	83.6	62.1	32.3	34.1	93.6	154.7	150.9
	1939	126.6	108.3	131.1	112.0	100.6	104.1	62.1	24.9	63.2	96.2	106.1	136.3
	1940	131.1	121.9	126.5	125.5	108.3	93.3	....	....	....	....	....	....

	Year	Jan.	Feb.	March	April	May	June	July	August	Sept.	Oct.	Nov.	Dec.
29 ε	1931	....	....	....	....	....	....	100.4	96.4	100.4	74.4	71.6	89.2
	1932	84.2	102.1	90.5	131.4	131.9	115.6	73.8	75.3	103.6	72.9	63.6	105.9
	1933	87.1	72.4	69.4	98.9	120.9	147.8	127.2	122.5	98.5	77.9	44.9	71.1
	1934	101.9	101.8	112.8	144.3	122.7	94.6	84.9	102.3	88.2	93.2	66.9	81.7
	1935	115.2	107.9	123.1	122.0	100.4	108.9	98.8	95.6	53.2	97.2	93.6	96.7
	1936	102.4	96.9	117.8	131.1	117.1	119.7	104.9	93.9	67.6	51.0	78.4	105.4
	1937	99.7	94.2	124.0	131.9	124.9	115.9	106.5	114.7	75.0	47.5	105.8	143.9
	1938	110.9	96.0	106.7	98.4	86.6	91.9	84.5	76.3	42.1	48.6	110.7	127.1
	1939	118.8	114.9	136.3	116.3	104.2	107.7	98.4	63.7	44.9	101.6	108.7	128.1
	1940	113.4	108.8	115.5	113.6	103.3	91.0	....	....	....	....	....	....

PROBLEMS BASED ON TABLE ε

	30	31	32 (a)	32 (b)	35	36
a	365,194	592,156	611,656.4	571,614.4	$\log^{-1}$ 5.72336	269,583.4
b	50,435.85	25,217.9	54,723.6	53,943.43	$\log^{-1}$ 0.04716	122,142.556
c	......	......	......	......	......	−7,967.344
			Trend Ordinates			
1931	365,194	365,194	......	355,840	355,000	269,583
1932	415,630	415,630	392,762	409,783	369,600	383,759
1933	466,066	466,066	447,485	463,727	435,500	481,999
1934	516,502	516,502	502,209	517,671	485,500	564,305
1935	566,937	566,938	556,932	571,614	528,900	630,676
1936	617,373	617,374	611,656	625,557	589,600	681,113
1937	667,809	667,809	666,380	679,501	657,160	733,614
1938	718,245	718,245	721,103	733,444	732,600	734,181
1939	768,681	768,681	775,827	787,387	816,600	736,814
1940	819,117	819,117	830,550	......	......	723,512

33 $\log a = 6.39018$ $\log b = 0.050566$

Year	Sales y	$\log y$	x	x^2	$x \log y$	$\log a + \log bx$	Approximate Anti-Logs
1932	1,135,491	6.05349	−4	16	−24,21396	6.18892	1,545,000
1933	1,573,512	6.13818	−3	9	−18.41454	6.23948	1,736,000
1934	2,177,919	6.33786	−2	4	−12.67572	6.29005	1,950,000
1935	3,252,244	6.49870	−1	1	− 6.49870	6.34061	2,191,000
1936	3,669,528	6.56458	0	0	0	6.39018	2,455,700
1937	3,915,889	6.59285	1	1	6.59285	6.44075	2,759,500
1938	2,000,985	6.30124	2	4	12.60248	6.49131	3,090,000
1939	2,866,796	6.45741	3	9	19.37223	6.54188	3,482,500
1940	3,692,328	6.56733	4	16	26.26932	6.59244	3,912,000
Totals	24,284,692	57.51164		60	3.03396		

36 * $a = 1{,}113{,}850$ $b = 624{,}964.6$ $c = -45.628.5$

Year	x	y	xy	x^2	x^2y	x^3	x^4	Ordinates of Trend
1931	0	1,973,090	0	0	0	0	0	1,113,850
1932	1	1,135,491	1,135,491	1	1,135,491	1	1	1,693,186
1933	2	1,573,512	3,147,024	4	6,294,048	8	16	2,181,265
1934	3	2,177,919	6,533,757	9	19,601,271	27	81	2,578,087
1935	4	3,252,244	13,008,976	16	52,035,904	64	256	2,883,552
1936	5	3,669,528	18,347,640	25	91,738,200	125	625	3,097,961
1937	6	3,915,889	23,495,334	36	140,972,004	216	1296	3,221,012
1938	7	2,000,985	14,006,895	49	98,048,265	343	2401	3,252,805
1939	8	2,866,796	22,934,368	64	183,474,944	512	4096	3,193,343
1940	9	3,692,328	33,230,952	81	290,078,568	729	6561	3,042,623
Totals	45	26,257,782	135,840,437	285	883,378,695	2025	15333	

* Table 22.

FUNCTIONAL RELATIONSHIPS

1 Solution depends upon figures taken

2 $a=0.15136$ $b=0.3821$

(a)

1935	4.69835
1936	4.73656
1937	4.46910
1938	3.47563
1939	3.47711
1940	3.85921
1941	5.27308

(b) 0.765

3

	y	x	y^2	x^2	xy	y_c	$y-y_c$	$(y-y_c)^2$
1927	132.3	95.4	17503.29	9101.16	12621.42	127.9	4.4	19.36
1928	130.8	96.7	17108.64	9350.89	12648.36	125.8	5.0	25.00
1929	132.5	95.3	17556.25	9082.09	12607.18	125.8	6.7	44.89
1930	126.0	86.4	15876.00	7464.96	10886.40	113.2	12.8	163.84
1931	103.9	73.0	10795.21	5329.00	7584.70	94.2	9.7	94.09
1932	86.5	64.8	7482.25	4199.04	5605.20	82.7	3.8	14.44
1933	84.1	65.9	7072.81	4342.81	5542.19	82.1	2.0	4.00
1934	93.7	74.9	8779.69	5610.01	7018.13	96.9	−3.2	10.24
1935	100.4	80.0	10080.16	6400.00	8032.00	104.1	−3.7	13.69
1936	101.3	80.8	10261.69	6528.64	8185.04	114.3	−13.0	169.00
1937	105.3	86.3	11088.09	7447.69	9087.39	105.3	0	0
1938	97.8	78.6	9564.84	6177.96	7696.08	102.1	−4.3	18.49
1939	95.2	77.1	9063.04	5944.41	7339.92	100.0	−4.8	23.04
1940	96.6	78.6	9331.56	6177.96	7592.76	102.1	−5.5	30.25
1941	105.5	87.3	11130.25	7708.84	9210.15	114.5	−9.0	81.00
	1591.9	1221.1	172693.77	100865.46	131656.92			711.33

$a=9.07$ $b=1.415$ $S_y=\sqrt{\frac{711.33}{15}}=\sqrt{47.42}=\pm 6.88$

4

	y	y^2	x	x^2	xy	y_c	$y-y_c$	$(y-y_c)^2$
1927	132.3	17503.29	139	19321	18389.7	121.7	10.6	112.36
1928	130.8	17108.64	149	22201	19489.2	126.8	4.0	16.00
1929	132.5	17556.25	146	21316	19345.0	125.2	7.3	53.29
1930	126.0	15876.00	126	15876	15876.0	135.1	−9.1	82.81
1931	103.9	10795.21	87	7569	9046.3	95.3	8.6	73.96
1932	86.5	7482.25	65	4225	5622.5	84.1	2.4	5.76
1933	84.1	7072.81	70	4900	5887.0	86.7	−2.6	6.76
1934	93.7	8779.69	90	8100	8433.0	96.8	3.1	9.61
1935	100.4	10080.16	108	11664	10843.2	105.9	−5.5	30.25
1936	101.3	10261.69	114	12996	11548.2	109.0	−7.7	59.29
1937	105.3	11088.09	121	14641	12741.3	112.5	−7.2	51.84
1938	97.8	9564.84	95	9025	9291.0	99.3	−1.5	2.25
1939	95.2	9063.04	93	8649	8853.6	98.3	−3.1	9.61
1940	96.6	9331.56	98	9604	9466.8	100.9	−4.3	18.49
1941	105.5	11130.25	122	14884	12871.0	113.1	−7.5	56.25
	1591.9	172693.77	1623	184791	177703.8			588.53

$a=51.11$ $b=0.50776$ $S_y=\sqrt{\frac{588.53}{15}}=\sqrt{39.235}=\pm 6.26$

5

	y	x	d_y	$(d_y)^2$	d_x	$(d_x)^2$	d_xd_y
1927	132.3	95.4	26.1	681.21	14.0	196.00	365.40
1928	130.8	96.7	24.7	610.09	15.3	234.09	377.91
1929	132.5	95.3	26.3	691.69	14.1	198.81	370.83
1930	126.0	86.4	19.9	396.01	5.0	25.00	99.50
1931	103.9	73.0	−2.2	4.84	−8.4	70.56	18.48
1932	86.5	64.8	−19.6	384.16	−16.6	275.56	325.36
1933	84.1	65.9	−22.0	48.40	−15.5	240.25	341.00
1934	93.7	74.9	−12.4	153.76	−6.5	42.25	80.60
1935	100.4	80.0	−5.7	32.49	−1.4	1.96	7.98
1936	101.3	80.8	−4.8	23.04	−0.6	0.36	2.88
1937	105.3	86.3	−0.8	0.64	4.9	24.01	−3.92
1938	97.8	78.6	−8.3	68.89	−2.8	7.84	23.24
1939	95.2	77.1	−10.9	118.81	−4.3	18.49	46.87
1940	96.6	78.2	−9.5	90.25	−3.2	10.24	30.40
1941	105.5	87.3	−0.6	0.36	5.9	34.81	−3.54
Totals	1591.9	1221.1		3304.64		1380.23	2082.99

$M_{A_y}=106.1$ $M_{A_x}=81.4$ $r=0.975$

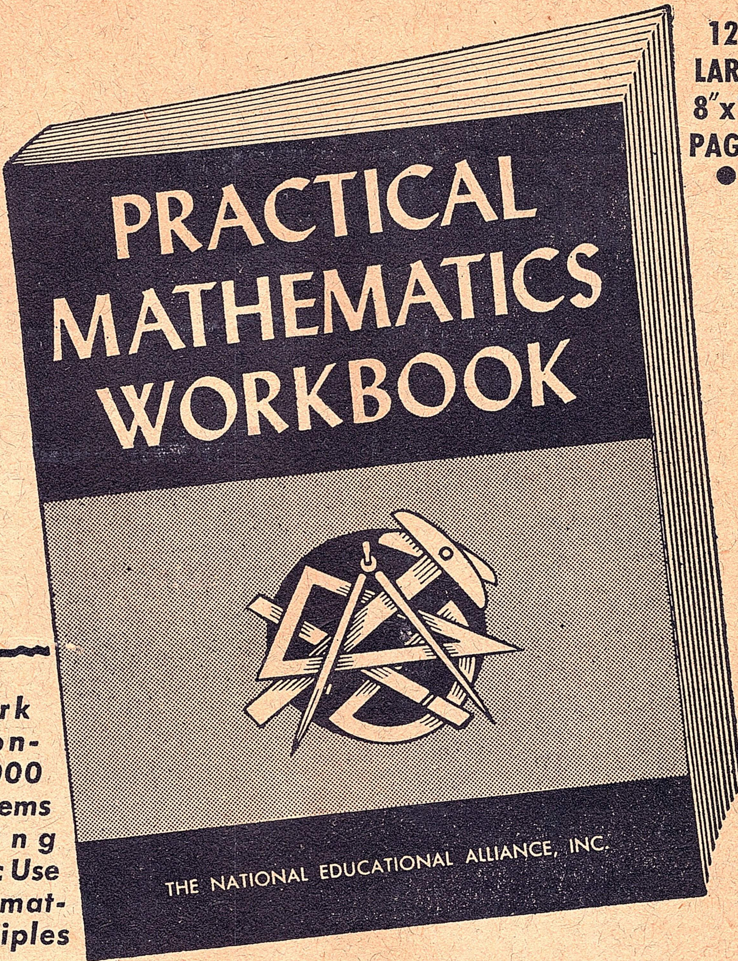
Index to Practical Mathematics

VOLUME III

only

$1

A Special Offer Made by Our **MEMBERS' SERVICE DEPARTMENT** exclusively to members now enrolled in the PRACTICAL MATHEMATICS Course



128 LARGE 8"x11" PAGES!

This Work Book Contains 1,000 Key Problems Covering Every Basic Use of Mathematical Principles

•

Enables You to Keep a Permanent Record of How You Arrived at Each Result

•

Invaluable for Ready Reference When You Encounter Similar Problems in Actual Work in Shop or Field

PRACTICAL MATHEMATICS

WORK BOOK

WOULDN'T you like to own this remarkably useful man-size WORK BOOK that brings you about 1,000 KEY PROBLEMS covering every basic principle in your Practical Mathematics Course—and provides ample space in which to do all your figuring and calculating in working out the answers? This is a *real convenience* that you will appreciate at every step of the Course. And it will have a *permanent value,* for it will provide for you in one compact place the actual methods of solution that you yourself have used in every fundamental type of problem. You can immediately refer to this record when you come across similar problems in the work you are now doing or may be called upon to do in the future.

SEE FURTHER DETAILS & ORDER FORM INSIDE THIS COVER